

Nanometer-MOSFETs für Digital- und Hochfrequenzanwendungen

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

vorgelegt der Fakultät für Elektrotechnik und Informationstechnik

der Technischen Universität Ilmenau

von

Dipl.-Ing. Ralf Granzner

Gutachter: (1) PD. Dr.-Ing. habil. Frank Schwier
(2) Prof. Dr.-Ing. Max Christian Lemme
(3) Univ.-Prof. Dr.-Ing. habil. Hannes Töpfer

Tag der Einreichung: 03.04.2012

Tag der wissenschaftlichen Aussprache: 11.03.2013

urn:nbn:de:gbv:ilm1-2013000097

Nanometer-MOSFETs für Digital- und Hochfrequenzanwendungen

Der Silizium-MOSFET (metal-oxide-semiconductor field-effect transistor) ist seit den 1980er Jahren das mit Abstand wichtigste Bauelement der Halbleiterelektronik. Geht man davon aus, dass der Trend der letzten Jahrzehnte beibehalten wird, ist um das Jahr 2020 mit der Massenproduktion von integrierten Schaltkreisen zu rechnen, deren Transistoren eine Gatelänge von nur noch 10nm besitzen. Bei so kleinen Abmessungen kann die Funktionsweise der Transistoren nur durch die Anwendung nichtklassischer MOSFET-Konzepte aufrechterhalten werden. In diesen Bauelementen muss mit einem starken Einfluss quasiballistischer Transporteffekte sowie mit einer Reihe quantenmechanischer Effekte gerechnet werden, über deren tatsächliche Auswirkungen auf die Performance der Transistoren und auf deren Empfindlichkeit gegenüber Prozessschwankungen noch keine endgültige Klarheit besteht.

Das Ziel der vorliegenden Arbeit ist es, verschiedene nichtklassische MOSFET-Konzepte hinsichtlich ihrer Performance sowie ihrer Eignung für zukünftige Logik- bzw. Analog/RF-Anwendungen zu bewerten und ihren Designspielraum einzugrenzen. Die Untersuchungen basieren auf numerischen Bauelementesimulationen unter Zuhilfenahme analytischer Berechnungen. Es wird gezeigt, dass das einfache und robuste Drift-Diffusionsmodell mit geringfügigen aber effektiven Modifikationen für die Simulation solcher dreidimensionalen Bauelemente geeignet ist.

Die Auswirkungen von Quanteneffekten auf die elektrischen Eigenschaften der Transistoren werden mit Hilfe eigens dafür entwickelter analytischer Modelle sowie mit numerischen Lösungen der Schrödinger- und Poissongleichungen untersucht.

Es wird gezeigt, dass bei einer Gatelänge von 10nm sehr verschiedene Querschnittsgeometrien zu ähnlichen elektrischen Parametern führen können. Sowohl Double-Gate, Tri-Gate aber auch Single-Gate SOI MOSFETs können die Anforderungen für High-Performance Logikanwendungen in diesem Gatelängenbereich erfüllen.

In einer detaillierten Untersuchung zum Einfluss von Geometrieschwankungen wird demonstriert, dass die Massenfertigung von MOSFETs mit 10nm Gatelänge extreme Anforderungen an die Prozesskontrolle stellt, insbesondere um Fluktuationen des Off-Stroms in vertretbaren Grenzen zu halten.

Die Simulation der Hochfrequenzeigenschaften dieser Transistoren liefert beeindruckende Grenzfrequenzen bis in den Terahertzbereich. Werden die untersuchten MOSFETs im Subthresholdbereich betrieben, erreichen sie bereits Transitfrequenzen im zweistelligen Gigahertzbereich. Stark skalierte nichtklassische MOSFETs haben demnach nicht nur das Potential für neue Anwendungen im Bereich des Terahertz-Gap sondern auch für Gigahertz-Elektronik mit extrem geringer Leistungsaufnahme.

Um das außerordentliche Potential dieser Transistoren auch tatsächlich nutzen zu können, ist die Minimierung äußerer Widerstände und Kapazitäten essentiell. Die Wirkung solcher parasitärer Elemente wird detailliert untersucht und die Obergrenzen für tolerierbare Widerstände und Kapazitäten werden bestimmt.

Nanometer MOSFETs for Digital and High-Frequency Applications

The Silicon MOSFET (metal-oxide-semiconductor field-effect transistor) is the dominating device in semiconductor electronics since the 1980s. Provided the trend of the past decades will continue, and many signs support this assumption, one can expect the mass production of integrated circuits (ICs) containing MOSFETs with gate lengths of only 10nm around the year 2020. At such small dimensions, transistor operation can only be maintained by using nonclassical MOSFET concepts. Moreover, in such devices a strong influence of quasi-ballistic carrier transport and quantum mechanical effects can be expected whose actual impact on the device performance and its immunity against process induced fluctuations is not finally clarified.

The aim of the present work is the evaluation of different nonclassical MOSFET concepts regarding their performance and suitability for future logic and analog/RF (radio frequency) applications and to define their design space. The investigations are based on numerical device simulations supported by analytical calculations. It is shown that the simple and robust drift-diffusion model with a few minor but effective modifications is well suited for the simulation of such three-dimensional devices.

The impact of quantum effects on the electrical device behavior is investigated by means of analytical models, developed in the frame of this work, and self-consistent numerical solutions of the Schrödinger and Poisson equations.

It can be shown that at the 10-nm gate length level, quite different channel geometries can lead to similar electrical parameters. Double-Gate, Tri-Gate and even Single-Gate SOI MOSFETs are able to meet the targets for high-performance logic applications.

A detailed study on the impact of geometry fluctuations on transistor performance demonstrates that the mass production of 10nm MOSFETs requires extreme process stability, in particular, to keep fluctuations of the off-current within tolerable limits.

The simulation of the high-frequency behavior of 10nm MOSFETs results in impressive frequency limits well within the terahertz range. In the subthreshold regime, these MOSFETs already show cutoff frequencies exceeding 50GHz. Hence, extremely scaled nonclassical MOSFETs have the potential not only for new applications within the so-called terahertz gap but also for ultra-low power RF electronics.

In order to really take advantage of the outstanding high frequency capabilities of these devices, minimizing extrinsic resistances and capacitances is essential. The effects of such parasitic elements are investigated in detail and upper limits for tolerable combinations of resistances and capacitances are elaborated.

Inhaltsverzeichnis

1. Einleitung	1
2. Grundlagen	6
2.1. Aufbau und Wirkungsweise des MOSFET	6
2.1.1. Wirkprinzip.....	6
2.1.2. MOSFET-Konzepte.....	8
2.2. Gleichstromverhalten	12
2.2.1. Kennlinien	12
2.2.2. Wichtige Kenngrößen.....	13
2.3. Hochfrequenzeigenschaften.....	15
2.3.1. Kleinsignalparameter.....	15
2.3.2. Grenzfrequenzen.....	17
2.4. Skalierung von MOSFETs in den Nanometerbereich	21
2.4.1. Kurzkanaleffekte	21
2.4.2. Konventionelle Skalierung	22
2.4.3. Probleme im Nanometerbereich	23
3. Simulation und Modellierung nanoskalierter MOSFETs	29
3.1. Simulationsverfahren und -modelle	29
3.1.1. Transportmodelle.....	30
3.1.2. Modellierung von Quantum Confinement.....	33
3.2. Simulation des Ladungstransports in Nanometer-MOSFETs	35
3.2.1. Vergleich der Transportmodelle.....	35

3.2.2. Anpassung des DDM und des HDM	39
3.3. Simulation der Hochfrequenzeigenschaften	45
3.3.1. Methoden der Kleinsignalanalyse	45
3.3.2. Vergleich der Transportmodelle	48
3.4. Quanteneffekte	53
3.4.1. Verschiebung der Schwellspannung	53
3.4.2. Einfluss von Quanteneffekten auf die Gatekapazität	60
3.4.3. Elektronenbeweglichkeit in Silizium-Nanodrähten	74
3.4.4. Korrektur klassischer Bauelementesimulationen	82
4. Nanometer-MOSFETs für digitale Logik	89
4.1. Einführung	86
4.2. Design von Single-Gate, Double-Gate und Tri-Gate MOSFETs	87
4.2.1. Simulationsmethode	87
4.2.2. Single-Gate und Double-Gate MOSFETs	90
4.2.3. Tri-Gate MOSFETs	94
4.2.4. Einfluss von Geometrieschwankungen	99
4.3. Einfluss von parasitären Widerständen	106
4.3.1. Theoretische Betrachtungen	107
4.3.2. Auswirkung von Serienwiderständen auf I_{on}	110
5. Nanometer-MOSFETs für Hochfrequenzanwendungen	112
5.1. Einführung	112
5.2. Design des inneren Transistors	114
5.2.1. Kanalquerschnitt	115
5.2.2. Äquivalente Oxiddicke	126

5.3. Einfluss parasitärer Elemente	132
5.3.1. Parasitäre Widerstände	132
5.3.2. Parasitäre Kapazitäten	137
6. Zusammenfassung und Ausblick	142
Anhang	147
A. Analytische Näherungslösungen für f_T und f_{\max}	147
B. Partielle Ableitungen für Gleichung (3-36).....	153
C. Quantenkapazität von Silizium-Nanodrähten mit Bulk-Eigenschaften.....	154
Literaturverzeichnis	156
Veröffentlichungen	166
Danksagung	169

Kapitel 1

Einleitung

Die rasante Entwicklung der Informationstechnologie der letzten Jahrzehnte ist im Wesentlichen auf die Massenproduktion immer komplexer werdender integrierter Schaltkreise (integrated circuit, IC) zurückzuführen. Diese exponentielle Entwicklung, welche seit der Erfindung des IC anhält, wurde im Jahr 1965 erstmals von Gordon E. Moore erkannt [1] und ist im Laufe der Zeit unter der Bezeichnung Moore'sches Gesetz weltberühmt geworden. In seiner damaligen Prognose, die sich nur auf die Daten einiger weniger damals existierender IC's stützen konnte, sagte Moore eine Verdopplung der Anzahl der Bauelemente pro Chip in regelmäßigen Zeitabständen voraus, ein Trend der bis heute andauert [2].

Der entscheidende Antrieb dieser Entwicklung ist bis in die Gegenwart die fortwährende Miniaturisierung der einzelnen Bauelemente (auch Skalierung genannt), wobei der Silizium (Si) MOSFET (metal-oxide-semiconductor field-effect transistor) seit den 1980er Jahren das mit Abstand wichtigste Bauelement in der Halbleiterelektronik ist [2]. Dieser Transistortyp ist der Grundbaustein der sogenannten CMOS-Logik (complementary MOS), auf der heutzutage nahezu alle digitalen Schaltungen basieren. Ein besonderer Vorteil des Si MOSFET liegt in seinem relativ einfachen Aufbau und der damit verbundenen guten Skalierbarkeit. So konnten Ingenieure anhand klar definierter Regeln [3] über mehr als zwei Jahrzehnte die Transistorabmessungen verkleinern ohne die Funktionsweise der Transistoren zu beeinträchtigen. Der Effekt der Skalierung war immer eine verbesserte Performance gegenüber der Vorgängergeneration, d. h. eine höhere Schaltgeschwindigkeit bei gleichzeitig geringerer Leistungsaufnahme pro Schaltoperation. Abgesehen von den immer gegenwärtigen technologischen Schwierigkeiten, war diese Art der Skalierung relativ unkompliziert und geradlinig und hat zu einem gesunden Wachstum der Halbleiterindustrie geführt, so dass das Moore'sche Gesetz im Laufe der Zeit zu einer sich selbst erfüllenden Prophezeiung geworden ist.

Seit einiger Zeit jedoch stößt die Skalierung des konventionellen Silizium-MOSFET an ihre Grenzen. So wurde es in den letzten Jahren zunehmend schwieriger, mit dem

konventionellen MOSFET-Design ein sicheres Abschalten der Transistoren zu gewährleisten. Aufgrund der geringen Bauelementeabmessungen und der damit verbundenen großen elektrischen Felder werden quantenmechanische Effekte relevant, die eine Verschlechterung der Transistoreigenschaften zur Folge haben. Dazu gehören Tunnelströme, welche die statische Verlustleistung erhöhen, oder die Quantisierung der Ladungsträgerenergie, wodurch die Herstellung identischer MOSFETs mit gleichen elektrischen Eigenschaften zunehmend schwieriger wird. Gegenwärtig sind die kritischen Transistorabmessungen, wie die Gatelänge oder die äquivalente Gateoxiddicke, in einen Bereich vorgestoßen, in dem der Übergang zu alternativen MOSFET Konzepten vollzogen werden muss, um weiterhin einen Performancegewinn durch die Skalierung der Transistoren zu ermöglichen.

Alternative bzw. nichtklassische Konzepte für den Silizium-MOSFET nutzen unter anderem die Vorteile von Silicon-on-Insulator (SOI)-Substraten oder basieren auf sogenannten Nanodrähten aus Silizium. Viele dieser Konzepte verwenden im Gegensatz zum konventionellen MOSFET mehr als nur ein Gate, man spricht von Multiple-Gate-Konzepten wie z.B. Double-Gate (DG), Tri-Gate (TG) oder Gate-All-Around (GAA). Die Struktur solcher Transistoren ist dreidimensional (3D) während der klassische MOSFET ein im Grunde zweidimensionales (2D) Bauelement ist. Darüber hinaus machen sich aufgrund der geringen Bauelementeabmessungen quantenmechanische Effekte und ballistische Transporeffekte immer stärker bemerkbar. Das alles führt dazu, dass die klassischen Regeln der Skalierung [3] auf diese modernen Transistortypen nicht mehr anwendbar sind. Die ITRS (International Technology Roadmap for Semiconductors), ein ständig aktualisiertes Standardwerk in der Halbleiterbranche, in dem wichtige Zielgrößen für zukünftige Technologiegenerationen definiert sind, macht zwar Vorgaben für die kritischen Abmessungen oder die elektrischen Parameter von Silizium MOSFETs bis zum Jahr 2024 [4], klare Richtlinien für das 3D Design nichtklassischer MOSFETs ergeben sich daraus allerdings nicht. Solche Angaben für die erforderliche Transistorgeometrie sind aber für die Technologieentwicklung von entscheidender Bedeutung.

Daraus ergibt sich die Motivation für die vorliegende Arbeit. Sie ist der Versuch, einen Designraum für stark skalierte nichtklassische MOSFETs, die die Vorgaben der ITRS theoretisch erfüllen können, zu definieren. Die Untersuchungen konzentrieren sich auf die MOSFET-Generation mit einer Gatelänge von 10nm, deren Einführung in die Massenproduktion von der ITRS um das Jahr 2020 erwartet wird [4].

Diese eigentlich recht einfache Aufgabenstellung entpuppt sich bei genauerer Betrachtung schnell als ein sehr vielschichtiges Problemfeld. Die bisher allgemein verwendeten Modelle und Simulationsverfahren versagen bei der Beschreibung dieser neuartigen Bauelemente mit Nanometerabmessungen. Die Schwierigkeiten beginnen bereits mit der dreidimensionalen Struktur der Transistoren. Will man solche Strukturen in numerischen Simulationen mit hinreichender Genauigkeit beschreiben, erhöht sich der rechentechnische Aufwand gegenüber den für die konventionellen MOSFETs ausreichenden zweidimensionalen Simulationen enorm. Im Prinzip lassen sich nichtklassische MOSFET-Strukturen zwar mit Hilfe analytischer Modelle ohne großen numerischen Aufwand beschreiben, solche Modelle existieren bisher allerdings nur in Ansätzen und beschreiben in der Regel immer nur ausgewählte elektrische Parameter stark vereinfachter Strukturen.

Die Probleme sind aber noch grundlegender. Eine ganze Reihe nichtklassischer physikalischer Effekte die bei MOSFET-Strukturen mit Nanometerdimensionen relevant sind, können mit den üblichen kommerziellen Simulatoren nicht oder nur teilweise und stark vereinfacht berücksichtigt werden. Es existieren zwar Simulationsmethoden, die verschiedene Aspekte des Ladungsträgertransports in solchen Bauelementen physikalisch korrekt beschreiben können. Diese sind jedoch in der Regel rechentechnisch sehr aufwendig und haben darüber hinaus immer das Problem, dass andere ebenfalls wichtige Effekte nur unzureichend oder gar nicht berücksichtigt werden. Das liegt zum Teil auch daran, dass bestimmte Aspekte des Elektronentransports, in Nanodrähten zum Beispiel, bisher noch nicht richtig verstanden sind. So ist die tatsächliche Auswirkung der Quantisierung der Elektronenenergie (Quantum Confinement) auf solche Parameter wie Beweglichkeit oder Sättigungsgeschwindigkeit noch nicht wirklich geklärt. Verschiedene theoretische und experimentelle Untersuchungen haben bisher zu kontroversen Ergebnissen geführt. Eine andere offene Frage ist beispielsweise die Auswirkung von Quanteneffekten auf die Gatekapazität nichtklassischer MOSFETs.

Kurz gesagt existiert derzeit noch kein theoretisches Verfahren, das alle relevanten Aspekte der Physik stark skaliert nichtklassischer MOSFETs berücksichtigt und mit dem die oben gestellte Aufgabe mit einem vertretbaren Aufwand zu lösen wäre. Die Entwicklung einer geeigneten Methode, mit der sich die Auswirkungen des 3D Designs auf die elektrischen Eigenschaften der betrachteten Transistoren simulieren lassen, war deshalb das erste wichtige Arbeitspaket im Rahmen dieser Dissertation. Dabei wurde ein ingenieurmäßiger Ansatz verfolgt. Das Ziel war nicht, alle auftretenden physikalischen

Effekte mit den bestmöglichen mathematischen Methoden korrekt zu beschreiben, sondern mit überschaubarem Aufwand die vorhandenen Modelle und Verfahren so zu erweitern, dass eine hinreichend genaue Beschreibung des Verhaltens von Nanometer-MOSFETs möglich wird.

Als Basiswerkzeug diene der kommerzielle Bauelementesimulator ATLAS [5] von Silvaco unter Verwendung des einfachen aber robusten Drift-Diffusionsmodells (DDM). Damit ist es möglich, wie im Rahmen dieser Dissertation gezeigt werden wird, mit einer einfachen Modifikation der Modellparameter die Effekte des nichtstationären Ladungstransports hinreichend genau zu simulieren. Die relevanten quantenmechanischen Effekte können ebenfalls mit einem relativ einfachen Ansatz berücksichtigt werden, nämlich durch nachfolgende Korrekturen der klassisch simulierten elektrischen Kenngrößen. Die dazu notwendigen Korrekturparameter lassen sich, abgekoppelt von den eigentlichen Bauelementesimulationen, durch Lösen der Schrödingergleichung in einer Schnittebene senkrecht zum Stromfluss bestimmen. Entsprechende Modelle wurden im Rahmen dieser Arbeit entwickelt.

Mit einem so konstruierten Verfahren ist es möglich, die elektrischen Kenngrößen für eine Vielzahl dreidimensionaler MOSFET-Strukturen zu simulieren. Im Rahmen dieser Arbeit wird eine Möglichkeit gezeigt, wie sich durch einen Vergleich mit ausgewählten Zielgrößen für zukünftige CMOS-Bauelemente der Designspielraum für die Querschnittsgeometrie nichtklassischer MOSFETs eingrenzen lässt. Darüber hinaus wird in einer detaillierten Studie die Auswirkung unvermeidbarer Fertigungstoleranzen der Kanalgeometrie auf eine Reihe wichtiger elektrischer Parameter untersucht. Solche Fluktuationen werden heute als ein wesentliches Hindernis für die MOSFET-Skalierung angesehen.

Ein weiterer Schwerpunkt der vorliegenden Dissertation liegt auf der Untersuchung der Hochfrequenzeigenschaften stark skalierten MOSFETs für analoge Anwendungen. Ausgangspunkt ist die Frage, ob die Zielvorgaben der ITRS für die Grenzfrequenzen f_T und f_{\max} tatsächlich erreicht werden können und wenn ja, unter welchen Voraussetzungen. Besonders interessant in diesem Zusammenhang ist der Einfluss unvermeidbarer parasitärer Widerstände und Kapazitäten. Ein Ergebnis dieser Untersuchungen ist die Bestimmung von Obergrenzen für die Kombination verschiedener parasitärer Elemente. Außerdem wird ein neuentwickeltes analytisches Kompaktmodell für f_{\max} vorgestellt, welches im Gegensatz zu den bereits existierenden Modellen den Einfluss des Drainwiderstandes richtig beschreibt.

Die vorliegende Arbeit ist folgendermaßen gegliedert. In Kapitel 2 werden die wesentlichen Grundlagen zum Verständnis moderner Si MOSFETs vorgestellt und die für die Arbeit relevanten elektrischen Größen definiert. Darüber hinaus werden entscheidende Probleme bei der MOSFET-Skalierung sowie verschiedene Lösungsansätze diskutiert.

Kapitel 3 befasst sich mit der Simulation und Modellierung stark skaliert MOSFETs. In einer Vergleichsstudie wird die Eignung verschiedener Transportmodelle des Bauelementesimulators ATLAS geprüft. Als Referenz werden dabei Monte-Carlo-Simulationen mit dem Simulator DAMOCLES [6] genutzt. Der Vergleich liefert unter anderem einen Satz modifizierter Modellparameter für das DDM, womit eine Simulation sowohl des Gleichstromverhaltens als auch der Hochfrequenzeigenschaften der betrachteten Transistoren möglich ist. Darüber hinaus werden die Auswirkungen von Quanteneffekten auf die elektrischen Eigenschaften nichtklassischer MOSFETs untersucht. Zu diesem Zweck werden analytische Modelle, mit denen sich der Einfluss des Quantum Confinement berechnen lässt entwickelt. Gesonderte Studien beschäftigen sich mit der Auswirkung von Quanteneffekten auf die Gatekapazität sowie auf die Elektronenbeweglichkeit in Tri-Gate MOSFETs. Schließlich werden Möglichkeiten zur Korrektur klassisch simulierter Kennlinien aufgezeigt.

In Kapitel 4 wird eine Möglichkeit demonstriert, den Designspielraum nichtklassischer MOSFETs für High-Performance CMOS-Anwendungen einzugrenzen. Darüber hinaus werden die Auswirkungen von Fluktuationen der Bauelementeabmessungen auf die elektrischen Parameter der Transistoren sowie der Einfluss parasitärer Widerstände diskutiert.

In Kapitel 5 werden die Hochfrequenzeigenschaften stark skaliert nichtklassischer MOSFETs untersucht, wobei ein besonderer Schwerpunkt auf der Wirkung parasitärer Widerstände und Kapazitäten liegt.

Den Abschluss bildet Kapitel 6 mit einer Zusammenfassung der wichtigsten Ergebnisse.

Kapitel 2

Grundlagen

In diesem Kapitel werden die wesentlichen Grundlagen zum Verständnis moderner MOSFETs vorgestellt. Dazu gehören Aufbau und Wirkungsweise, Gleichstrom- und Hochfrequenzeigenschaften sowie einige Aspekte der Skalierung von MOSFETs. Sofern nicht anders angegeben, beziehen sich alle Diskussionen auf n-Kanal MOSFETs (nMOSFETs), gelten aber zumindest qualitativ auch für p-Kanal MOSFETs (pMOSFETs).

2.1. Aufbau und Wirkungsweise des MOSFET

2.1.1. Wirkprinzip

Wie jeder Transistor, so ist auch der MOSFET im Prinzip ein mikro- bzw. nanoelektronischer Schalter. Seine Schalterfunktion wird durch ein sehr einfaches Wirkprinzip realisiert: Zwischen zwei Gebieten mit hoher Leitfähigkeit, genannt Source und Drain, befindet sich ein Kanalgebiet, dessen Leitfähigkeit in einem weiten Bereich durch die am Gate angelegte Spannung gesteuert werden kann. Der prinzipielle Aufbau eines MOSFET ist in Abbildung 2.1 dargestellt. Source und Drain sind typischerweise hoch dotierte Halbleitergebiete und stellen Reservoirs an beweglichen Ladungsträgern dar. Im nMOSFET sind sie n-dotiert und die beweglichen Ladungsträger sind Elektronen. Im pMOSFET sind sie p-dotiert und die beweglichen Ladungsträger sind Löcher. Das Gate besteht aus Metall oder aus hoch dotiertem Polysilizium und ist über einen Isolator kapazitiv an das Kanalgebiet gekoppelt. Der traditionelle Gateisolator im Si MOSFET ist Siliziumoxid (SiO_2), wobei neuerdings auch Hafnium-basierte Isolatoren mit hoher Dielektrizitätskonstante zum Einsatz kommen. Der Kanal befindet sich in einem Halbleitergebiet, welches wesentlich geringer dotiert ist als die Source- und Draingebiete. Außerdem hat das Kanalgebiet gewöhnlich einen anderen Leitungstyp als Source und Drain, ist also im nMOSFET p-dotiert. Die Schichtfolge

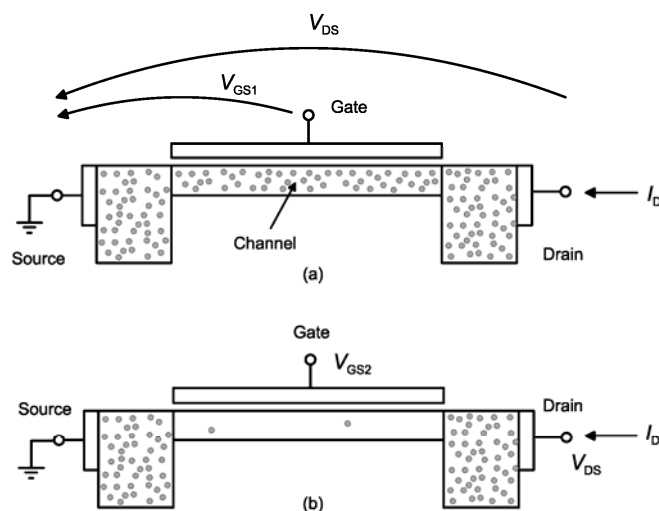


Abbildung 2.1 Prinzipielle Struktur eines nMOSFET [2]. (a) On-Zustand: die angelegte Gatespannung V_{GS1} ist größer als die Schwellspannung V_{th} . (b) Off-Zustand: V_{GS2} ist kleiner als V_{th} .

Metall – Oxid – Halbleiter (metal – oxide – semiconductor, MOS) unterhalb des Gates stellt einen Kondensator dar und gibt dem MOSFET seinen Namen.

Das Potential am Source-Kontakt ist das Referenzpotential im Transistor und liegt gewöhnlich auf Masse (0V). Wird nun eine ausreichend hohe positive Spannung V_{GS} an das Gate eines nMOSFET angelegt, so induziert das elektrische Feld in der MOS-Kapazität (daher der Name Feldeffekttransistor) eine große Zahl von Elektronen im Kanal. So entsteht eine gut leitende Verbindung zwischen Source und Drain und bei angelegter Drain-Source-Spannung V_{DS} kann ein Drainstrom I_D fließen. Der Transistor ist eingeschaltet bzw. befindet sich im On-Zustand.

Ist die Gatespannung V_{GS} kleiner als die so genannte Schwellspannung V_{th} , befinden sich nur sehr wenige Elektronen im Kanal und die Verbindung zwischen Source und Drain hat eine sehr geringe Leitfähigkeit. Bei angelegter Drain-Source-Spannung kann praktisch kein Drainstrom mehr fließen, d.h. der MOSFET ist ausgeschaltet bzw. befindet sich im Off-Zustand. Ist der MOSFET bei $V_{GS} = 0$ ausgeschaltet ($V_{th} > 0$), spricht man vom Anreicherungstyp (enhancement type bzw. normally-off). Ist er dagegen bei $V_{GS} = 0$ eingeschaltet ($V_{th} < 0$), handelt es sich um den Verarmungstyp (depletion type bzw. normally-on). Für CMOS-Schaltungen verwendet man MOSFETs vom Anreicherungstyp.

Heutzutage gibt es eine große Zahl verschiedener MOSFET-Konzepte, die sich im Design stark unterscheiden können. Sie alle basieren aber auf der prinzipiellen Struktur

aus Abbildung 2.1. Im nachfolgenden Abschnitt werden die wichtigsten MOSFET-Varianten kurz vorgestellt.

2.1.2. MOSFET-Konzepte

A. Der konventionelle Bulk-MOSFET

Der konventionelle Bulk-MOSFET stellt die ursprüngliche Variante des MOSFET dar und ist auch gegenwärtig noch der mit Abstand meistverwendete Transistor in der Digitaltechnik. Die Struktur eines Bulk-nMOSFET ist in Abbildung 2.2 skizziert. In ein einkristallines p- dotiertes Siliziumsubstrat, auch als Bulk bezeichnet, sind die hoch dotierten n^+ -Source- und Draingebiete eingebracht. Das „+“ kennzeichnet eine sehr hohe Konzentration von Dotanten (hier: Donatoren). Für die ohmschen Kontakte an Source und Drain werden Metall-Legierungen, in modernen MOSFETs vor allem aber Silizide verwendet. Das Gate besteht meist aus n^+ -Polysilizium und der Gateisolator aus SiO_2 , dem so genannten Gateoxid. Polysilizium ist, wie der Name schon sagt, kein Einkristall, sondern polykristallines Silizium. Der ursprüngliche Grund für die Einführung von Polysiliziumgates waren die qualitativ besseren und zuverlässigeren Gateoxide, die sich damit im Vergleich zu Metall-Gates herstellen ließen. Darüber hinaus lässt sich Polysilizium bei entsprechender Dotierung sowohl für n- als auch für pMOSFETs als Gatematerial einsetzen, ein entscheidender Vorteil für CMOS-Technologien.

Das Si-Substrat kann über den Bulk-Kontakt auf ein definiertes Potential gelegt werden,

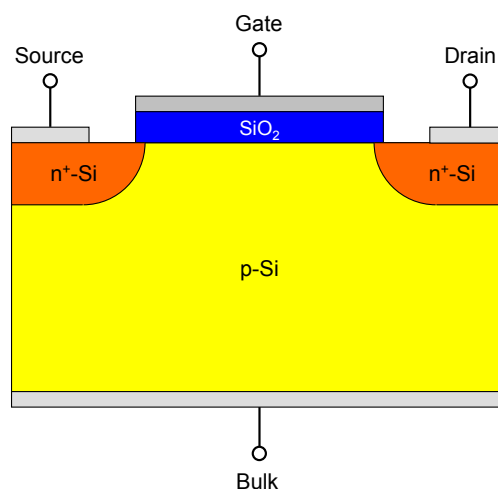


Abbildung 2.2 Schematischer Querschnitt eines Bulk-nMOSFET.

normalerweise auf Source-Potential (Masse). Mit der Gatespannung lässt sich nun das Potential sowie die Elektronen- und Löcherkonzentrationen an der Oberfläche des p-Substrats (d.h. an der Grenzfläche zum Gateoxid) einstellen. Beim Anreicherungstyp ist bei $V_{GS} = 0$ die Oberfläche mit Löchern angereichert oder bereits leicht verarmt. Letzteres ist normalerweise in nMOSFETs mit n⁺-Polysiliziumgates der Fall. Eine positive Gatespannung hat zur Folge, dass die positiv geladenen Löcher von der Oberfläche zurückgedrängt werden. Es entsteht eine Verarmungszone mit einer negativen Raumladung, die von den negativ geladenen Akzeptor-Ionen (Dotanten im p-Gebiet) gebildet wird. Gleichzeitig werden die negativ geladenen Elektronen zur Oberfläche hingezogen. Ab einer bestimmten positiven Gatespannung ist die Elektronendichte an der Oberfläche größer als die Löcherdichte, d.h. der Leitungstyp kehrt sich um. Man spricht dann von Inversion. Der Inversionskanal ist nur wenige Nanometer dick und dünner als die Verarmungszone. Die maximale Ausdehnung der Verarmungszone hängt von der Dotierung ab und liegt im Bereich von einigen zehn bis einigen hundert Nanometern. Der Großteil des Substrats bleibt also unbeeinflusst von der Gatespannung. Damit der Transistor bei $V_{GS} = 0$ auch tatsächlich abschaltet, darf kein Stromfluss zwischen Source und Drain im Inneren des Substrats möglich sein. Dies wird mit der p-Dotierung des Substrats und der daraus resultierenden n⁺pn⁺-Anordnung zwischen Source und Drain sichergestellt. Bei jeder angelegten Drain-Source-Spannung ist einer der beiden pn-Übergänge in Sperrichtung gepolt. Wichtig ist eine ausreichend hohe p-Dotierung, so dass die Raumladungszonen der beiden pn-Übergänge sich nicht gegenseitig beeinflussen. Dies ist mit zunehmender Skalierung immer schwieriger zu realisieren. Darüber hinaus machen sich bei hohen Substratdotierungen quantenmechanische Effekte wie Band-zu-Band-Tunneln bemerkbar, die zu einer Erhöhung des Leckstroms im ausgeschalteten Zustand beitragen. Der Bulk-MOSFET ist daher wahrscheinlich nicht bis zu extrem kleinen Gatelängen um 10nm skalierbar [4].

Gegenwärtig haben die kleinsten kommerziellen Bulk-MOSFETs Gatelängen um 30nm, und kommen beispielsweise in den Intel® Core™ Prozessoren zum Einsatz [7].

B. SOI MOSFETs

Silicon-on-Insulator (SOI) MOSFETs haben Vorteile gegenüber dem konventionellen Bulk-MOSFET bezüglich der Skalierbarkeit. Der prinzipielle Aufbau eines SOI MOSFET ist in Abbildung 2.3 dargestellt. Die MOSFET-Strukturen sind hier in einer dünnen einkristallinen Siliziumschicht realisiert, die sich auf einer so genannten vergrabenen SiO₂-Schicht (buried oxide, BOX) befindet. Je nach Dicke der Si-Schicht

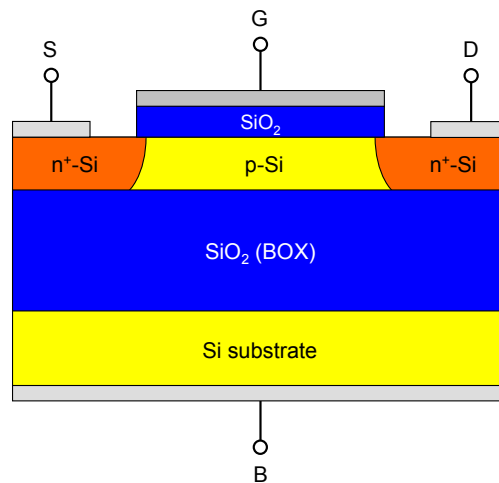


Abbildung 2.3 Schematischer Querschnitt eines SOI MOSFET.

unterscheidet man zwischen teilweise verarmten (partially depleted, PD) und vollständig verarmten (fully depleted, FD) SOI MOSFETs. In PD SOI MOSFETs ist die Si-Schicht dicker als die maximale Ausdehnung der Verarmungszone. Es existiert also unterhalb der Verarmungszone ein vom Gate nicht kontrollierter Substratbereich, so dass hier im Prinzip die gleichen Forderungen bezüglich der Substratdotierung gelten wie beim Bulk-MOSFET. In FD SOI MOSFETs ist die Si-Schicht dünner als die Verarmungszone, so dass das gesamte Gebiet zwischen Source und Drain vom Gate kontrolliert werden kann. Bei sehr dünnen Si-Schichten ist keine p-Dotierung zum Unterdrücken von Leckströmen erforderlich und der Kanal kann auch undotiert bleiben. Man bezeichnet diese Transistoren als ultra-thin body (UTB) SOI MOSFETs. Das FD SOI Konzept ist also grundsätzlich besser für eine Skalierung bis zu extrem kleinen Gatelängen geeignet als der Bulk-MOSFET. Weitere Vorteile gegenüber dem Bulk-MOSFET sind geringere parasitäre Kapazitäten sowie die bessere Isolierung gegenüber benachbarten Bauelementen.

Die ersten kommerziellen SOI MOS-Schaltkreise kamen 1998 auf den Markt [2]. Seither wächst die Zahl der Anwendungen der SOI-Technologie ständig. Gegenwärtig kommen SOI MOSFETs zum Beispiel in AMD's aktuellen Athlon™ und Phenom™ Prozessoren zum Einsatz [8].

C. Multiple-Gate MOSFETs

Die konsequente Weiterentwicklung des FD SOI Konzepts führte zu MOSFETs mit mehr als nur einem Gate, den so genannten Multiple-Gate (MG) MOSFETs. Im

Gegensatz zum gewöhnlichen planaren SOI MOSFET mit einem Gate an der Oberfläche, auch Single-Gate (SG) SOI MOSFET genannt, wird das Kanalgebiet in MG MOSFETs von verschiedenen Seiten aus gesteuert. Dadurch kann die Potentialbarriere zwischen Source und Drain effektiver mit der Gatespannung kontrolliert werden. Deshalb eignen sich MG MOSFETs besonders gut für eine Skalierung bis in den Bereich von 10nm Gatelänge und weniger [4].

Die Bauformen von MG MOSFETs sind vielfältig. Gegenwärtig am bedeutendsten sind MG Konzepte mit rechteckigem Querschnitt wie sie in Abbildung 2.4 skizziert sind. Dazu gehören Double-Gate (DG) MOSFETs mit zwei gegenüberliegenden Gates, Tri-Gate (TG) MOSFETs mit Gates an drei Seiten sowie Gate-All-Around (GAA) MOSFETs bei denen der Kanal vollständig vom Gate umschlossen ist. Bei DG MOSFETs unterscheidet man je nach Lage der Gates zwischen planaren und vertikalen DG MOSFETs. Letztere bezeichnet man auch als FinFETs. Darüber hinaus gibt es MG Designs mit dreieckigem oder auch rundem Kanalquerschnitt [9]-[10].

Verschiedene Gruppen haben bereits experimentelle DG MOSFETs [11]-[15], Tri-Gate MOSFETs [16]-[21] und auch GAA MOSFETs [22]-[27] erfolgreich hergestellt. Zwar gibt es noch keine kommerzielle Anwendung von MG MOSFETs, jedoch hat Intel im Mai 2011 angekündigt, die Massenproduktion von Tri-Gate MOSFETs mit 22nm Gatelänge bis Ende 2011 vorzubereiten [28].

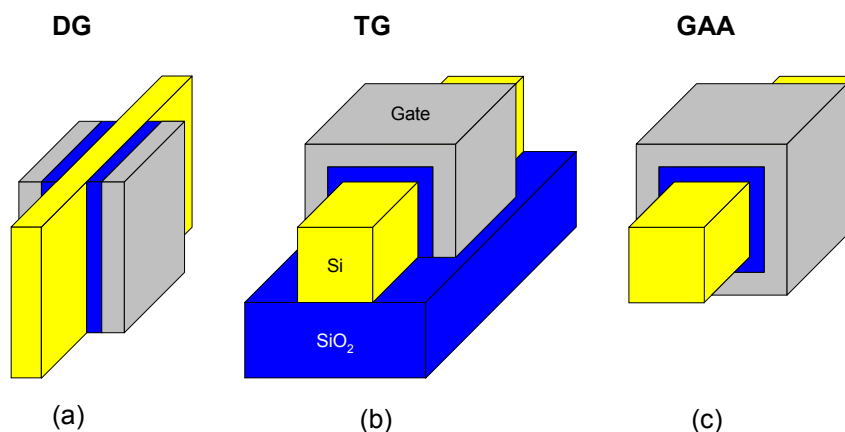


Abbildung 2.4 Schematischer Aufbau verschiedener Multiple-Gate MOSFETs. (a) Double-Gate MOSFET, (b) Tri-Gate MOSFET (c) Gate-All-Around MOSFET.

2.2. Gleichstromverhalten

2.2.1. Kennlinien

In allen gut konstruierten MOSFETs haben die Gleichstromkennlinien den gleichen prinzipiellen Verlauf. Man unterscheidet Ausgangskennlinien $I_D(V_{DS})$ und Transferkennlinien (Übertragungskennlinien) $I_D(V_{GS})$. Eine Eingangskennlinie wie beim Bipolartransistor gibt es nicht, da der Eingangsstrom, also der Gatestrom I_G , praktisch Null ist. Die Kennlinien von n-Kanal und p-Kanal MOSFETs unterscheiden sich qualitativ nur im Vorzeichen der Spannungen und Ströme (positiv im nMOSFET, negativ im pMOSFET). Abbildung 2.5 zeigt das Ausgangskennlinienfeld eines nMOSFET. Die einzelnen Kennlinien $I_D(V_{DS})$ unterscheiden sich aufgrund der unterschiedlichen Gatespannung V_{GS} . Die Kennlinien lassen sich in zwei Bereiche unterteilen, in den aktiven oder linearen Bereich und den Pinch-Off- oder Sättigungsbereich. Im Sättigungsbereich ist der Drainstrom nahezu unabhängig von V_{DS} und wird vor allem von V_{GS} bestimmt. Die Grenze zwischen aktivem Bereich und Sättigungsbereich befindet sich bei einer Drain-Source-Spannung von $V_{DS} = V_{GS} - V_{th}$.

In Abbildung 2.6 ist eine typische Transferkennlinie im linearen und im logarithmischen Maßstab dargestellt. Im linearen Maßstab wird deutlich, dass erst ab $V_{GS} > V_{th}$ ein merklicher Drainstrom fließt. Der logarithmische Maßstab zeigt darüber hinaus, dass auch bei $V_{GS} < V_{th}$, im so genannten Subthresholdbereich, der Drainstrom nicht null ist. Anhand der Transferkennlinie lassen sich die wichtigsten Kenngrößen des MOSFET, welche in dieser Arbeit betrachtet werden, erläutern.

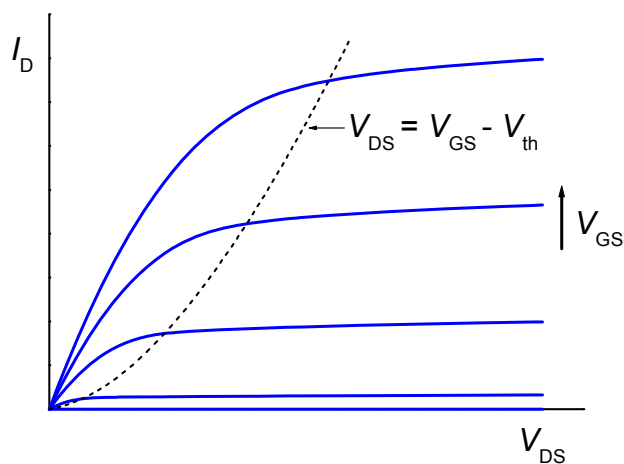


Abbildung 2.5 Ausgangskennlinienfeld eines nMOSFET. Die gestrichelte Linie kennzeichnet die Grenze zwischen aktivem Bereich und Sättigungsbereich.

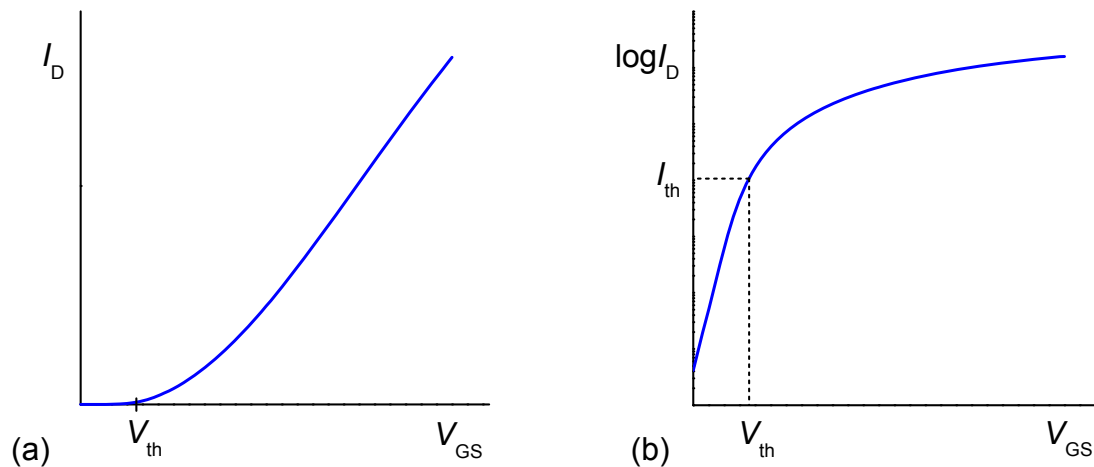


Abbildung 2.6 Transferkennlinie eines nMOSFET im (a) linearen und (b) logarithmischen Maßstab.

2.2.2. Wichtige Kenngrößen

Die wichtigsten Kenngrößen des MOSFET lassen sich besonders anschaulich anhand der Transferkennlinie erläutern.

Schwellspannung

Als Schwellspannung V_{th} wird die Gatespannung bezeichnet, bei der gerade ein leitfähiger Kanal gebildet wird. Sie stellt die Grenze zwischen Subthresholdbereich und On-Zustand dar. Obwohl es sich bei V_{th} um eine eher willkürliche Größe handelt, ist sie ein wichtiger Parameter in den meisten analytischen MOSFET-Modellen und muss beim Design von Schaltungen berücksichtigt werden. In den folgenden Untersuchungen wird eine messtechnische Definition der Schwellspannung verwendet. Hier ist V_{th} die Gatespannung, bei der ein bestimmter Drainstrom $I_D = I_{th}$ fließt. Es gilt

$$I_{th} = 10^{-7} \text{ A} \cdot \frac{w_{eff}}{L_G}, \quad (2-1)$$

wobei w_{eff} die effektive Gateweite und L_G die Gatelänge des Transistors bezeichnet. Die Definition von w_{eff} hängt vom Transistortyp ab. In planaren (Single-Gate) MOSFETs ist w_{eff} gleich der tatsächlichen Gateweite W_G . In MG MOSFETs tragen zu w_{eff} die Weiten aller den Kanal steuernden Gates bei. So ist z.B. in DG MOSFETs $w_{eff} = 2W_G$.

Der Strom I_{th} ist proportional zu einer bestimmten Elektronenschichtkonzentration $n_S = n_{sth}$ im Kanal von nMOSFETs. Typische Werte für n_{sth} liegen im Bereich von 10^{10} cm^{-2} . Ein n_{sth} -Kriterium eignet sich besonders für die Bestimmung der Schwellspannung von MOS-Strukturen ohne Source und Drain.

Subthreshold Slope

Mit Subthreshold Slope S wird der inverse Anstieg der logarithmischen Transferkennlinie im Subthresholdbereich bezeichnet.

$$S = \left(\frac{d(\lg I_D)}{dV_{GS}} \right)^{-1} \quad (2-2)$$

Er beschreibt die Güte des Abschaltverhaltens und sollte so klein wie möglich sein. Für MOSFETs mit Langkanaleigenschaften kann S mit

$$S = \frac{k_B T}{q} \ln 10 \left(1 + \frac{C_{depl}}{C_{ox}} \right) \quad (2-3)$$

abgeschätzt werden [29]. Dabei bezeichnet C_{depl} die Verarmungskapazität, C_{ox} die Oxidkapazität, k_B die Boltzmannkonstante, T die Temperatur und q die Elementarladung. Der Ausdruck $k_B T/q$ wird auch Temperaturspannung V_T genannt und beträgt bei Zimmertemperatur ($T = 300\text{K}$) etwa 26mV. Im Bulk-MOSFET ist C_{depl} die Kapazität der Verarmungszone unterhalb des Gates, in FD SOI MOSFETs die Kapazität der Reihenschaltung aus Verarmungskapazität der Si-Schicht und der Kapazität des BOX. In gut designten FD SOI MOSFETs ist $C_{depl} \ll C_{ox}$, so dass die theoretische untere Grenze von $S \approx 60\text{mV/dec}$ erreicht wird (d.h. eine Änderung der Gatespannung um 60mV bewirkt eine Änderung des Drainstroms um eine Dekade).

Steilheit

Mit Steilheit (transconductance) g_m bezeichnet man den Anstieg der Transferkennlinie im linearen Maßstab.

$$g_m = \left. \frac{dI_D}{dV_{GS}} \right|_{V_{DS}=\text{const.}} \quad (2-4)$$

Sie ist ein Maß für die Gate-Steuerwirkung im On-Zustand und sollte so groß wie möglich sein.

On-Strom und Off-Strom

Der On-Strom I_{on} und der Off-Strom I_{off} sind wahrscheinlich die beiden wichtigsten Kenngrößen für MOSFETs in digitalen Schaltungen. Sie sind folgendermaßen definiert

$$I_{\text{on}} = I_{\text{D}}(V_{\text{GS}} = V_{\text{DS}} = V_{\text{DD}}), \quad I_{\text{off}} = I_{\text{D}}(V_{\text{GS}} = 0, V_{\text{DS}} = V_{\text{DD}}), \quad (2-5)$$

wobei V_{DD} die Betriebsspannung der Schaltung ist. Mit Hilfe dieser beiden Ströme lässt sich eine Reihe wichtiger Eigenschaften des einzelnen Transistors sowie der gesamten Schaltung abschätzen. Zum Beispiel geht I_{on} in die so genannte CV/I-Metrik ein, ein Maß für die intrinsische Verzögerungszeit τ des MOSFET [30].

$$\tau \approx \frac{C \cdot V_{\text{DD}}}{I_{\text{on}}} \quad (2-6)$$

Dabei gibt $C = C_{\text{G}} + C_{\text{par}}$ die gesamte Gatekapazität des Transistors an, inklusive der intrinsischen Gatekapazität C_{G} sowie parasitärer Überlappungs- und Streukapazitäten C_{par} . Für ein geringes τ sollte I_{on} möglichst groß sein. Der Reziprokwert $1/\tau$ ist ein Maß für die maximale Schaltfrequenz des MOSFET.

Der Off-Strom ist für die statische Verlustleistung des MOSFET,

$$P_{\text{stat}} = I_{\text{off}} \cdot V_{\text{DD}}, \quad (2-7)$$

verantwortlich und sollte daher möglichst klein sein.

2.3. Hochfrequenzeigenschaften

Neben den typischen Anwendungen in der Digitaltechnik spielen MOSFETs in analogen Hochfrequenzschaltungen eine immer wichtigere Rolle. Zur Charakterisierung von Transistoren für solche Anwendungen werden üblicherweise Kenngrößen verwendet, die aus dem Kleinsignalverhalten der Bauelemente abgeleitet sind. In den folgenden Abschnitten werden die zum Verständnis notwendigen theoretischen Grundlagen erläutert und die für die Arbeit wichtigen Kenngrößen vorgestellt.

2.3.1. Kleinsignalparameter

Wird ein Transistor in einem bestimmten Gleichstromarbeitspunkt betrieben, so kann die Wirkung kleiner Änderungen um den Arbeitspunkt durch einen Satz von

Kleinsignalparametern vollständig beschrieben werden. Zu diesem Zweck wird der Transistor als Vierpol (auch Zweitor, two-port) betrachtet, eine Black Box, deren Klemmverhalten durch eine Reihe von Vierpolgleichungen gegeben ist. Eine typische Vierpolschaltung ist in Abbildung 2.7 zu sehen. Mit i_1 und v_1 werden Eingangstrom und Eingangsspannung des Vierpols bezeichnet, mit i_2 und v_2 Ausgangstrom und Ausgangsspannung. Da es sich hierbei um sinusförmige Wechselgrößen handelt, werden kleine Buchstaben verwendet.

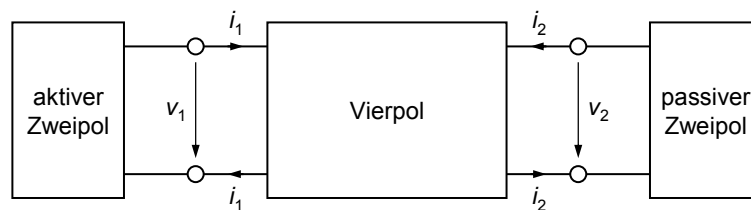


Abbildung 2.7 Prinzipielle Vierpolschaltung.

Generell kann jeder Vierpol durch verschiedene äquivalente Sätze von Kleinsignalparametern beschrieben werden, beispielsweise durch Y-, H-, Z- oder A-Parameter. Für Transistoren verwendet man gewöhnlich die so genannten Leitwert- oder Y-Parameter oder auch die Hybrid- oder H-Parameter. Die dazugehörigen Vierpolgleichungen lauten

$$\begin{aligned} i_1 &= y_{11}v_1 + y_{12}v_2 \\ i_2 &= y_{21}v_1 + y_{22}v_2 \end{aligned} \quad (2-8)$$

bei Verwendung von Y-Parametern sowie

$$\begin{aligned} v_1 &= h_{11}i_1 + h_{12}v_2 \\ i_2 &= h_{21}i_1 + h_{22}v_2 \end{aligned} \quad (2-9)$$

bei Verwendung von H-Parametern. Alle Größen in (2-8) und (2-9), die Kleinsignalparameter sowie die Spannungen und Ströme sind komplex und haben im Allgemeinen frequenzabhängige Beträge. Aus den Vierpolgleichungen lassen sich die Definitionen der Kleinsignalparameter ableiten. Bei wechselstrommäßigem Kurzschluss am Eingang ($v_1 = 0$) bzw. am Ausgang ($v_2 = 0$) folgen aus (2-8) die Definitionsgleichungen der Y-Parameter

$$\begin{aligned}
 y_{11} &= \left. \frac{i_1}{v_1} \right|_{v_2=0}, & y_{12} &= \left. \frac{i_1}{v_2} \right|_{v_1=0}, \\
 y_{21} &= \left. \frac{i_2}{v_1} \right|_{v_2=0}, & y_{22} &= \left. \frac{i_2}{v_2} \right|_{v_1=0}.
 \end{aligned}
 \tag{2-10}$$

2.3.2. Grenzfrequenzen

Mit Hilfe der Kleinsignalparameter lassen sich wichtige Transistorkenngrößen definieren, zum Beispiel die Kurzschlussstromverstärkung

$$h_{21} = \frac{y_{21}}{y_{11}}, \tag{2-11}$$

oder die unilaterale Leistungsverstärkung nach Mason [31]

$$U = \frac{|y_{21} - y_{12}|^2}{4[Re(y_{11})Re(y_{22}) - Re(y_{12})Re(y_{21})]}. \tag{2-12}$$

Beide Größen hängen stark von der Frequenz f ab. Bei hohen Frequenzen zeigen sie einen typischen Abfall von etwa 20dB/dec (Abbildung 2.8). Die Einheit Dezibel (dB) ist für h_{21} und U unterschiedlich definiert, nämlich

$$h_{21}[\text{dB}] = 20 \lg |h_{21}| \tag{2-13}$$

sowie

$$U[\text{dB}] = 10 \lg |U|. \tag{2-14}$$

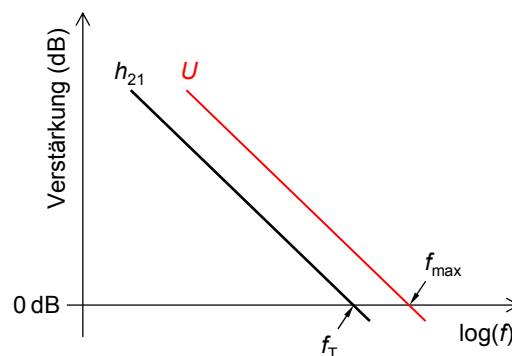


Abbildung 2.8 Typische Frequenzabhängigkeiten von h_{21} und U .

Entsprechend Abbildung 2.8 lassen sich für jeden Transistor zwei charakteristische Frequenzen angeben, die Grenzfrequenzen f_T und f_{\max} . Die Transitfrequenz f_T ist definiert als die Frequenz, bei der der Betrag von h_{21} gleich 1 (bzw. 0dB) wird. Das bedeutet, bei Frequenzen größer als f_T kann der Transistor den Eingangsstrom nicht mehr verstärken. Die maximale Schwingfrequenz f_{\max} ist die Frequenz, bei der U gleich 1 (bzw. 0dB) wird. Dementsprechend findet bei Frequenzen oberhalb von f_{\max} keine Leistungsverstärkung mehr statt. Bei der Bestimmung von f_T und f_{\max} macht man sich das typische Frequenzverhalten von h_{21} und U zunutze. Für eine Messfrequenz f_m , die viel kleiner als f_T oder f_{\max} sein kann, werden h_{21} und U bestimmt. Durch Extrapolation mit -20dB/dec lassen sich daraus die Grenzfrequenzen f_T und f_{\max} ermitteln. Hierbei gilt

$$f_T = f_m \cdot |h_{21}| \quad (2-15)$$

und

$$f_{\max} = f_m \cdot \sqrt{|U|}. \quad (2-16)$$

Ein praktisches Hilfsmittel zur Abschätzung von f_T und f_{\max} stellen analytische Näherungslösungen dar, die sich auf ein bestimmtes Kleinsignalersatzschaltbild des betrachteten Transistors beziehen. Abbildung 2.9 zeigt ein für MOSFETs häufig verwendetes Kleinsignalersatzschaltbild [32]-[34]. Es beinhaltet alle wichtigen intrinsischen und parasitären Elemente eines MOSFET mit Ausnahme von Leitungsinduktivitäten. Ein Bulk-Anschluss wird nicht explizit berücksichtigt, da er die Ersatzschaltung unnötig verkomplizieren würde und seine Wirkung mit den vorhandenen Elementen ausreichend gut beschrieben werden kann. Die Bezeichnungen der Elemente sind in Tabelle 2.1 zusammengefasst.

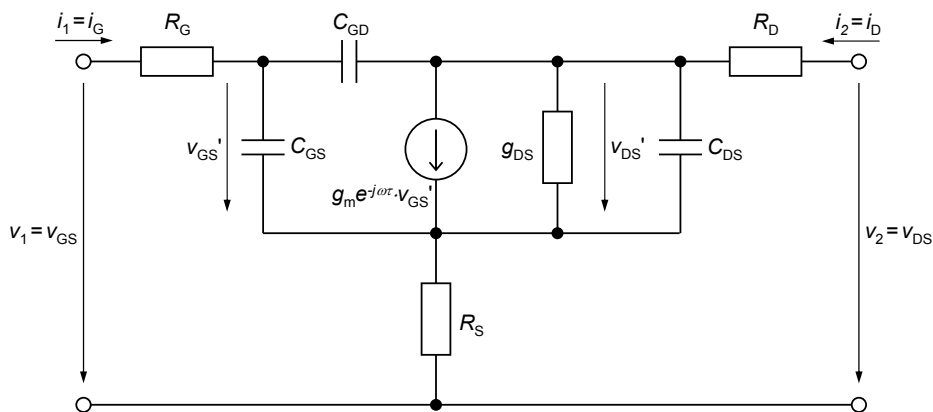


Abbildung 2.9 Kleinsignalersatzschaltbild für einen MOSFET mit parasitären Elementen.

g_m	Steilheit des inneren Transistors
g_{DS}	Drainleitwert des inneren Transistors
C_{GS}	Gate-Source Kapazität
C_{GD}	Gate-Drain Kapazität
C_{DS}	Drain-Source Kapazität
R_G	Gatewiderstand
R_S	Sourcewiderstand
R_D	Drainwiderstand

Tabelle 2.1 Bezeichnung der Elemente des Ersatzschaltbildes in Abbildung 2.9.

Die spannungsgesteuerte Stromquelle im Ersatzschaltbild ist im Wesentlichen durch die Steilheit g_m des inneren Transistors¹ definiert. Bei sehr hohen Frequenzen kann jedoch der Drainstrom Änderungen der Gatespannung nicht mehr unmittelbar folgen, was mit dem Faktor $e^{j\omega\tau}$ berücksichtigt wird. Dabei ist $\omega = 2\pi f$ die Kreisfrequenz, j die imaginäre Einheit und τ eine Zeitkonstante, die mit der Dauer des Elektronentransports vom Source-Ende zum Drain-Ende des Kanals zusammenhängt [35].

Der Drainleitwert ist definiert als

$$g_{DS} = \left. \frac{dI_D}{dV_{DS}} \right|_{V_{GS}=\text{const.}} \quad (2-17)$$

Die Kapazitäten C_{GS} und C_{GD} lassen sich sehr anschaulich mit Hilfe der Ladung auf der Gateelektrode Q_{Gate} definieren,

$$C_{GS} = \left. \frac{dQ_{\text{Gate}}}{dV_{GS}} \right|_{V_{GD}=\text{const.}} \quad (2-18)$$

und

$$C_{GD} = \left. \frac{dQ_{\text{Gate}}}{dV_{GD}} \right|_{V_{GS}=\text{const.}}, \quad (2-19)$$

¹ Als innerer Transistor wird hier der Transistor ohne externe parasitäre Elemente, d.h. ohne R_G , R_S und R_D bezeichnet.

wobei V_{GD} die Spannung zwischen Gate und Drain bezeichnet. Die Spannungen in (2-17) bis (2-19) sind in Übereinstimmung mit Abb. 2.9 die Spannungen am inneren Transistor.

Die Drain-Source-Kapazität C_{DS} kann man als die Kapazität des Substrats interpretieren [36]. Sie beschreibt die kapazitive Kopplung von Source und Drain. In SOI MOSFETs z.B. geschieht das zu einem wesentlichen Teil über das vergrabene Oxid.

Analysiert man das Ersatzschaltbild in Abbildung 2.9 unter Berücksichtigung der Gleichungen (2-10) – (2-12) sowie (2-15) und (2-16), lassen sich f_T und f_{max} mit Hilfe der Ersatzelemente ausdrücken. Man kann beispielsweise zeigen, dass im Fall $C_{DS} \ll C_{GS}, C_{GD}$ näherungsweise gilt

$$f_T = \frac{g_m}{2\pi(C_{GS} + C_{GD}) \left[1 + g_{DS}(R_S + R_D) + \frac{C_{GD}}{C_{GS} + C_{GD}} g_m(R_S + R_D) \right]}, \quad (2-20)$$

$$f_{max} = \frac{f_T}{2 \left[\frac{g_{DS}(R_G + R_S)}{N} + 2\pi f_T C_{GD} \left(R_G + \frac{C_{GD}}{C_{GS} + C_{GD}} R_D \right) \right]^{1/2}} \quad (2-21)$$

wobei

$$N = 1 + g_{DS}(R_S + R_D) + g_m R_S.$$

Mehr Details zur Herleitung und zur Genauigkeit von (2-20) und (2-21) sind in Anhang A zu finden. Gleichung (2-20) entspricht exakt dem bekannten Ausdruck für f_T von Tasker und Hughes [37]. Im Grenzfall vernachlässigbarer R_S und R_D vereinfacht sich (2-20) zu der wohl am häufigsten verwendeten Formel für f_T ,

$$f_T = \frac{g_m}{2\pi(C_{GS} + C_{GD})}. \quad (2-22)$$

Im Unterschied zu (2-20) und (2-22) ist Gleichung (2-21) ein neuer Ausdruck für f_{max} , der in der vorliegenden Arbeit erstmals vorgestellt wird. Der Vorteil von (2-21) gegenüber ähnlichen Gleichungen anderer Autoren [35], [38]-[40] besteht darin, dass in (2-21) auch der Einfluss von R_D auf f_{max} berücksichtigt wird.

Die Verwendung von (2-20) und (2-21) ist dann sinnvoll, wenn die Ersatzelemente bekannt sind oder sich relativ leicht bestimmen lassen. Darüber hinaus bieten sie die Möglichkeit, Trends abzuschätzen bzw. Trends gemessener oder simulierter Grenzfrequenzen zu interpretieren.

2.4. Skalierung von MOSFETs in den Nanometerbereich

2.4.1. Kurzkanaleffekte

Die Verkleinerung der Bauelementedimensionen ist seit Erfindung des Integrierten Schaltkreises (integrated circuit, IC) die wichtigste Voraussetzung für die ständige Erhöhung der Komplexität von ICs. Außer der Verringerung des Platzbedarfs pro Transistor soll mit der Skalierung vor allem die Schaltgeschwindigkeit $1/\tau$ erhöht werden. Besonders wichtig dafür ist eine Verkleinerung der Gatelänge. Die bloße Reduzierung von L_G hat jedoch eine Reihe unerwünschter Kurzkanaleffekte zur Folge, von denen die beiden wichtigsten im Folgenden kurz erläutert werden sollen.

A. Schwellspannungsabfall (Threshold Voltage Roll-Off)

Der auffälligste Effekt bei einer Kanallängenverkürzung ist die Verringerung der Schwellspannung, auch V_{th} -Roll-Off genannt. Dieser Effekt wird häufig als *der* Kurzkanaleffekt angesehen. Er lässt sich sehr anschaulich mit dem Modell der Ladungsteilung (charge sharing) nach Yau [41] beschreiben. In Abbildung 2.10 sind die Vermungszonen in einem Lankanal- und einem Kurzkanaltransistor skizziert. Das Kanalgebiet wird durch das Gatefeld verarmt, zu einem gewissen Anteil aber auch durch die Raumladungszonen der n^+p -Übergängen an Source und Drain. Dieser Anteil ist umso größer, je kürzer der Kanal ist. Das führt dazu, dass bei Erhöhung von V_{GS} die Inversion im Kurzkanaltransistor früher einsetzt als im Langkanaltransistor. Die Schwellspannung des Kurzkanaltransistors ist also kleiner. Die Schwellspannung hängt umso stärker von L_G ab, je kürzer der Kanal ist. Dieser Effekt ist deshalb so kritisch für integrierte Schaltungen, da dadurch Fertigungstoleranzen verstärkt zu unterschiedlichen V_{th} nominell gleichartiger MOSFETs führen können, was sich negativ auf

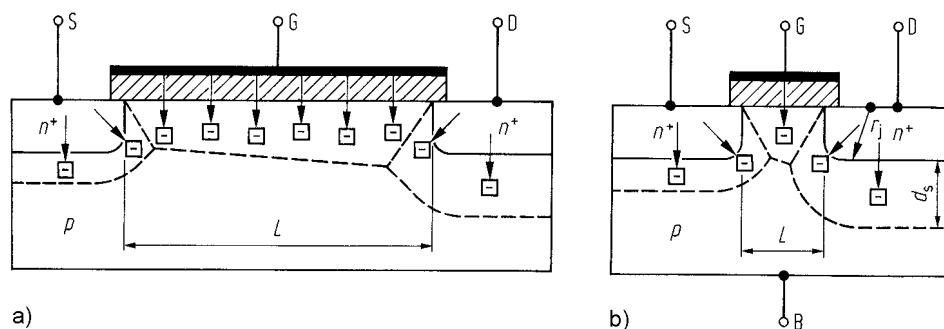


Abbildung 2.10 Raumladungszonen im Bulk-MOSFET [42]. (a) Langkanaltransistor. (b) Kurzkanaltransistor.

Geschwindigkeit und Verlustleistung, eventuell auch auf die Funktionsfähigkeit des IC auswirkt.

B. DIBL

Bei sehr kurzen Kanälen wird die Höhe der Potentialbarriere zwischen Source und Drain nicht mehr ausschließlich von der Gatespannung bestimmt, sondern kann auch durch eine positive Drain-Source-Spannung verringert werden. Dieser Effekt wird Drain-Induced Barrier Lowering, kurz DIBL, genannt [43]. In Abbildung 2.11 ist der Einfluss von V_{DS} auf die Barrierenhöhe im Langkanal- und im Kurzkanaltransistor skizziert. Der DIBL-Effekt bewirkt in erster Linie eine Verringerung der Schwellspannung mit steigendem V_{DS} . Dies wird gewöhnlich mit dem Parameter *DIBL* quantifiziert.

$$DIBL = \frac{\Delta V_{th}}{\Delta V_{DS}} \quad (2-23)$$

Die Einheit von *DIBL* ist mV/V. Ein *DIBL* von -15mV/V bedeutet demnach, dass sich die Schwellspannung des MOSFET um 15mV verringert, wenn V_{DS} um 1V erhöht wird.

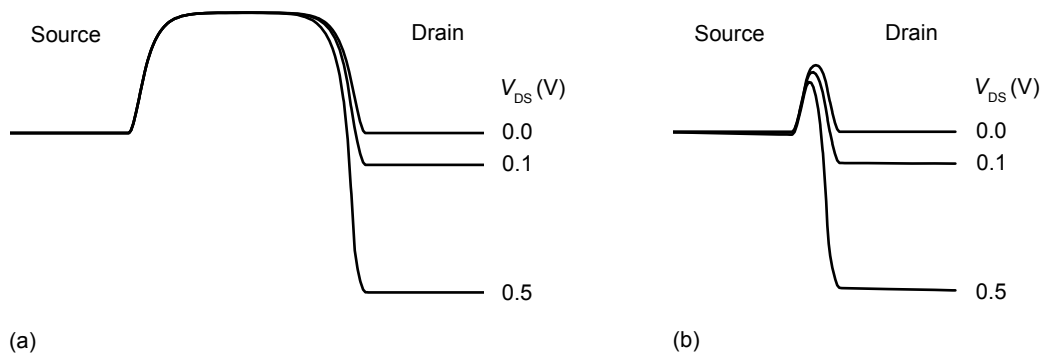


Abbildung 2.11 Leitbandkante im nMOSFET zwischen Source und Drain im Off-Zustand. (a) Langkanaltransistor. (b) Kurzkanaltransistor.

2.4.2. Konventionelle Skalierung

Um zu verhindern, dass bei einer Verkleinerung der MOSFET-Abmessungen die oben beschriebenen Kurzkanaleffekte das Bauelementeverhalten beeinträchtigen, wurde eine Reihe von grundlegenden Prinzipien der Skalierung entwickelt. Die drei wichtigsten Konzepte heißen Konstantfeldskalierung [3], allgemeine Skalierung [44] und

Konstantspannungsskalierung [29]. Das Prinzip bei der Konstantfeldskalierung beruht auf der Idee, alle lateralen und vertikalen Abmessungen sowie die angelegten Spannungen um den gleichen Faktor zu reduzieren, so dass die elektrischen Felder im Bauelement unverändert bleiben. Dieses Prinzip berücksichtigt jedoch nicht, dass die Halbleitereigenschaften wie z.B. die Bandlücke (bandgap) E_G unverändert bleiben, so dass die Schwellspannung nicht im gleichen Maße reduziert wird [29]. Aus diesem Grund werden bei der allgemeinen Skalierung die Spannungen weniger stark skaliert als die Abmessungen bzw. bleiben bei der Konstantspannungsskalierung unverändert. Das hat jedoch höhere Leistungsdichten und Feldstärken im Transistor zur Folge. Ein weiteres grundlegendes Problem ist die Nichtskalierbarkeit des Subthresholdverhaltens. Nach Gleichung (2-3) kann S einen Minimalwert, welcher nur von der Temperaturspannung abhängt, nicht unterschreiten. Eine Verringerung von V_{th} hat daher grundsätzlich einen höheren Off-Strom zur Folge.

Trotz der angesprochenen Probleme basierte die Skalierung des Bulk-MOSFET in der Vergangenheit im Wesentlichen auf diesen drei Konzepten. Die Miniaturisierung des Bulk-MOSFET stößt jedoch in Bereichen von einigen 10nm Gatelänge an fundamentale Grenzen. Die auftretenden Probleme sowie Möglichkeiten zu deren Lösung sollen im Folgenden kurz erläutert werden.

2.4.3. Probleme im Nanometerbereich

Band-zu-Band Tunneln

Bei der Skalierung des Bulk-MOSFET ist es notwendig, die Substratdotierung zu erhöhen. Für Gatelängen im Bereich von wenigen 10nm sind sehr hohe Dotierungen in der Größenordnung von 10^{19}cm^{-3} notwendig, so dass besonders am drainseitigen pn-Übergang sehr hohe Feldstärken auftreten. Dadurch kommt es verstärkt zu quantenmechanischem Band-zu-Band Tunneln, wodurch der Off-Strom erhöht wird [45]. Dieses Problem kann im Prinzip nur durch den Übergang zu FD SOI MOSFETs mit geringer Kanaldotierung umgangen werden.

Dotierungsfluktuationen

Bereits in MOSFETs mit $L_G \approx 100\text{nm}$ liegt die Anzahl der Dotierungsatome in der vom Gate kontrollierten Verarmungszone in der Größenordnung von einigen Hundert. Das bedeutet, die Kanaldotierung kann nicht mehr als ein Kontinuum von definierter Dichte

angesehen werden, sondern viel eher als eine Ansammlung von einzelnen Atomen. In Simulationen [46] konnte gezeigt werden, dass bei solchen Dimensionen nicht nur die absolute Anzahl sondern auch die mikroskopische Verteilung der Dotanten einen bedeutenden Einfluss auf die Schwellspannung des Transistors hat. Da die Zahl von Dotierungsatomen in einem Volumenelement des Kanalgebiets zufällig ist, kann auch die Schwellspannung von MOSFETs mit gleicher integraler Kanaldotierung gewissen Schwankungen unterliegen. Abbildung 2.12 zeigt simulierte Transferkennlinien [46] von 24 MOSFETs mit unterschiedlicher Dotantenverteilung bei gleicher integraler Dotierung ($N_A = 8.6 \cdot 10^{17} \text{cm}^{-3}$). Es ist zu sehen, dass die Kennlinien auf der Spannungsachse gegeneinander verschoben sind. Die mittlere Schwankung von V_{th} beträgt hier etwa 20mV. Diese Schwankungen werden durch Quanteneffekte noch verstärkt und sind umso größer, je kleiner die Kanalabmessungen sind [47]. Für MOSFETs mit einer effektiven Kannallänge von 30nm wurde in Simulationen eine V_{th} -Unbestimtheit von 100mV ermittelt [48]. Das ist für CMOS-Anwendungen zu viel.

Einen Ausweg bieten hier FD SOI Konzepte, sowohl SG als auch MG MOSFETs, mit undotiertem Kanal. Bei solchen MOSFETs besteht jedoch nicht mehr die Möglichkeit, V_{th} über die Kanaldotierung zu definieren. Die Schwellspannung kann dann nur noch über die Austrittsarbeit des Gatematerials eingestellt werden. Für CMOS-Anwendungen werden sogenannte Midgap-Materialien benötigt [2]. Das sind Metalle deren Fermikante nahe der Mitte der Silizium-Bandlücke liegt, zum Beispiel Wolfram (W), Titanitrid (TiN) oder auch Kobaltsilizid (CoSi_2) mit einer Austrittsarbeit Φ_M im Bereich von 4.6eV [49].

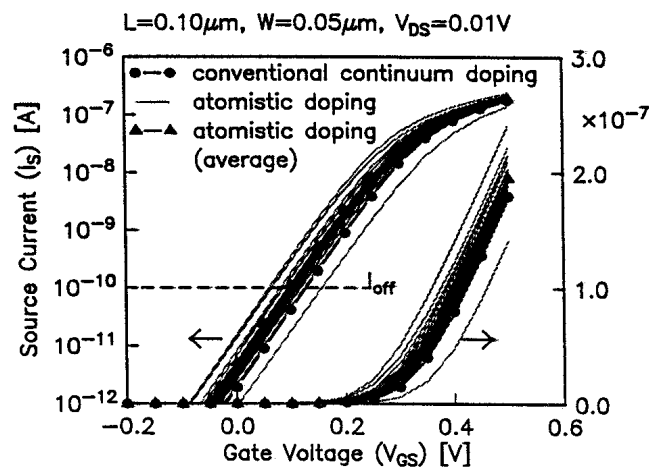


Abbildung 2.12 Simulierte Transferkennlinien (hier Sourcestrom I_S anstelle I_D) bei zufälliger Verteilung der Dotanten. Die integrale Kanaldotierung ist bei allen Transistoren gleich [46].

Verarmung im Polysilizium-Gate (Polydepletion)

Die Verwendung von Polysilizium als Gatematerial führt in extrem skalierten MOSFETs zu einem weiteren Problem. Durch das elektrische Feld im Gatedielektrikum wird nicht nur das Kanalgebiet verarmt sondern auch das Polysilizium-Gate an der Grenze zum Gateoxid [50]-[51]. Die Verarmungszone im Polysilizium wirkt wie ein zusätzliches Dielektrikum und die tatsächliche Gatekapazität verringert sich dadurch. Das hat negative Auswirkungen auf das Abschaltverhalten und den On-Strom. Dieser Effekt ist umso größer, je dünner das Gatedielektrikum ist. Bei Oxidicken von $t_{\text{ox}} \approx 1\text{nm}$ ist er nicht mehr zu vernachlässigen. Der Ausweg ist auch hier die Verwendung von Metallen als Gatematerial.

Gateoxid

Bei sehr dünnen Oxiden oder sehr hohen Feldstärken im Gateoxid ist die Potentialbarriere für Elektronen so schmal, dass sie hindurch tunneln können. Es entsteht ein unerwünschter Gatestrom, der in hoch integrierten Schaltungen einen erheblichen Beitrag zur statischen Verlustleistung liefern kann. Ab Oxiddicken $t_{\text{ox}} < 2\text{nm}$ werden die Tunnelströme zu groß für viele Anwendungen [52]. MOSFETs mit Gatelängen im 10nm-Bereich benötigen jedoch wesentlich dünnere Oxide [4].

Ein Ausweg ist die Verwendung von Isolatoren mit einer größeren Dielektrizitätskonstante (DEK) als die von SiO_2 , den so genannten high-K Materialien. Diese können bei gleicher Gatekapazität physikalisch dicker gemacht werden, was die Tunnelströme reduziert. Die äquivalente Oxiddicke EOT eines solchen Dielektrikums, d.h. die Dicke, die ein Gateisolator aus SiO_2 für die gleiche Gatekapazität haben müsste, ergibt sich aus

$$EOT = \frac{\epsilon_{r,\text{SiO}_2}}{\epsilon_r} \cdot t_d. \quad (2-24)$$

Dabei ist t_d die Dicke des Dielektrikums, ϵ_r ist dessen relative DEK und $\epsilon_{r,\text{SiO}_2}$ die relative DEK von SiO_2 .

Verschiedene high-K Materialien kommen infrage. Naheliegend ist zum Beispiel die Verwendung von Siliziumnitrid (Si_3N_4) mit $\epsilon_r \approx 7.5$ (zum Vergleich: $\epsilon_{r,\text{SiO}_2} \approx 3.9$). Es wird in der Siliziumtechnologie häufig verwendet, beispielsweise als Maskenmaterial. Für die Anforderungen in extrem skalierten MOSFETs ist seine DEK aber noch zu gering.

Gegenwärtig konzentriert sich der Großteil der Forschungs- und Entwicklungsarbeiten auf high-K Dielektrika mit einer relativen DEK von mehr als 20, wie z.B. Hafniumoxid (HfO_2) oder auch Zirkoniumoxid (ZrO_2), beide mit $\epsilon_r \approx 25$ [53]. Die größten Probleme liegen derzeit noch in der Zuverlässigkeit solcher Dielektrika, der thermischen Stabilität sowie unerwünschter Effekte wie Hysteresen oder der Reduzierung der Elektronenbeweglichkeit [54]-[56].

Letzteres ist eine direkte Folge der höheren dielektrischen Polarisierung in high-K Materialien im Vergleich zu SiO_2 . Die Dipole im Dielektrikum sind die Quelle optischer Phononen, an welchen Elektronen im Kanal gestreut werden können [57]. Eine höhere Zahl von Streuprozessen macht sich makroskopisch als eine Verringerung der Elektronenbeweglichkeit bemerkbar.

Ein Ausweg ist auch hier die Verwendung eines Metallgate [58]-[59]. Durch die extrem hohe Elektronendichte im Metall werden die Dipole im Dielektrikum abgeschirmt und die Einkopplung von optischen Phononen in den Kanal verringert sich deutlich.

Im Jahr 2007 ging bei Intel der erste Prozessor mit high-K Dielektrikum und Metallgates in Massenproduktion [59].

Direktes Tunneln von Source nach Drain

Ein weiterer Tunneleffekt kann das elektrische Verhalten extrem skaliert MOSFETs beeinträchtigen. Bei entsprechend kleiner Kanallänge ist die Potentialbarriere zwischen Source und Drain so schmal, dass die Elektronen direkt hindurch tunneln können (SD-Tunneln). Dadurch wird der Off-Strom erhöht und S steigt. Anders als Band-zu-Band-Tunneln ist SD-Tunneln unabhängig von der Kanaldotierung und stellt ein fundamentales Limit für die Skalierung von MOSFETs dar. Theoretische Arbeiten haben jedoch gezeigt, dass dieser Effekt erst ab Gatelängen jenseits 10nm von Bedeutung ist [60]-[63]. In [60] und [62] wird die ultimative Grenze bei $L_G \approx 5\text{nm}$ angegeben.

Ausbildung von Subbändern (Quantum Confinement)

Die Elektronen im Inversionskanal eines MOSFET bewegen sich in einem schmalen Potentialgraben. Ihre Bewegungsfreiheit ist in einer oder auch zwei Dimensionen eingeschränkt, was sich in einer Quantisierung des Leitungsbandes im Potentialgraben

bemerkbar macht. Man spricht von Quantum Confinement². Es kommt zur Ausbildung von Subbändern, deren Grundzustandsenergie oberhalb der Leitbandkante liegt [29].

Im Bulk-MOSFET wird der Potentialgraben durch das große senkrechte Feld an der Oberfläche verursacht, was gleichbedeutend ist mit einer starken Bandverbiegung [Abb. 2.13(a)]. In fully depleted SG oder MG MOSFETs stellt die Siliziumschicht selbst bereits einen Potentialgraben für Elektronen dar [Abb. 2.13(b)]. Die Wirkung des Quantum Confinement ist hier umso stärker, je kleiner der Kanalquerschnitt ist.

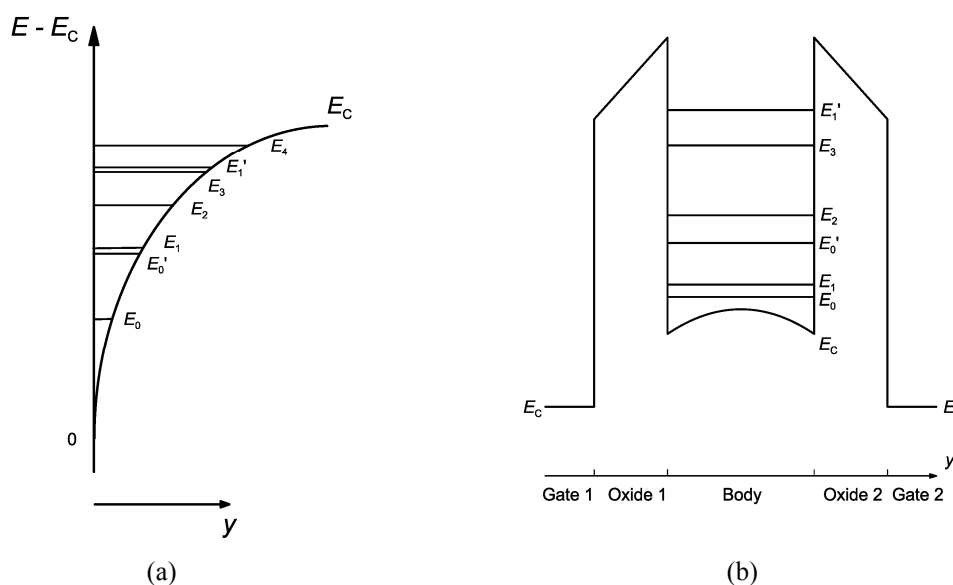


Abbildung 2.13 Potentialgraben und Subbandenergien im (a) Bulk-MOSFET und (b) Double-Gate MOSFET [2].

Die Ausbildung von Subbändern hat im Wesentlichen zwei Auswirkungen auf das elektrische Verhalten. Einerseits ist im Vergleich zur klassischen Erwartung die Schwellspannung größer, andererseits die Steilheit geringer. Der zweite Effekt führt zu einer Reduzierung des On-Stroms. Für die Skalierung ist jedoch der erste Effekt problematischer, weil dadurch V_{th} -Schwankungen infolge von Technologietoleranzen verstärkt werden. Im Falle des Bulk-MOSFET spielen die bereits erwähnten Dotierungsfluktuationen eine Rolle. Besonders kritisch bei FD SOI MOSFETs sind Toleranzen der Querschnittsgeometrie. Bei Abmessungen von wenigen Nanometern werden die Quanteneffekte so stark, dass bereits Abweichungen im Bereich von

² Eine wirklich gute deutsche Übersetzung des englischen Begriffs Quantum Confinement gibt es nicht. Eine mögliche Variante wäre zum Beispiel „quantenmechanische (Größen-)Beschränkung (freier Elektronen...)“.

Atomlagen zu erheblichen Änderungen der Schwellspannung führen [64]-[66].

Inwiefern dieser Effekt eine fundamentale Grenze der Skalierung darstellt, hängt also wesentlich von der technologischen Entwicklung ab. Wenn es gelingt, Strukturbreiten und Schichtdicken auf die Atomlage genau in Massenproduktion zu fertigen, lassen sich die Auswirkungen des Quantum Confinement beherrschen.

Fazit

Zukünftige extrem skalierte MOSFETs werden standardmäßig Metall-Gates und Gateisolatoren aus high-K Material besitzen. Der Bulk-MOSFET wird mit großer Wahrscheinlichkeit nicht bis zu Gatelängen um 10nm skaliert werden können. Die ultimativen MOSFETs basieren höchstwahrscheinlich auf FD SOI-Konzepten (SG, MG) mit undotiertem Kanal. Deshalb werden in den folgenden Untersuchungen ausschließlich solche MOSFET-Typen betrachtet.

Kapitel 3

Simulation und Modellierung nanoskalierter MOSFETs

Grundlage dieser Arbeit ist die numerische Bauelementesimulation. Dieses Hilfsmittel ermöglicht es, eine Vielzahl unterschiedlicher Transistorstrukturen in einem weiten Designbereich mit angemessener physikalischer Genauigkeit zu beschreiben. Eine umfassende Berücksichtigung aller auftretenden physikalischen Effekte ist jedoch, wenn überhaupt, nur mit extrem großen Rechenaufwand möglich. Besonders problematisch ist beispielsweise die gleichzeitige und selbstkonsistente Beschreibung von nichtstationärem Ladungstransport einerseits und Quantum Confinement andererseits. Deshalb wurden im Rahmen dieser Arbeit bestimmte Effekte mit Hilfe analytischer Modelle untersucht.

Nach einem kurzen Überblick über die derzeit relevanten Simulationsverfahren werden in den folgenden Abschnitten die in dieser Arbeit verwendeten Verfahren und Modelle vorgestellt und deren Auswahl begründet. Darüber hinaus werden verschiedene Aspekte des Ladungstransports sowie die Wirkung des Quantum Confinement in Nanometer-MOSFETs genauer untersucht und diskutiert.

3.1. Simulationsverfahren und -modelle

Die Mehrzahl der heute üblichen numerischen Simulationsverfahren basiert auf einer semiklassischen Theorie der Halbleiter. Die quantenmechanische Beschreibung des Halbleiterkristalls liefert die grundlegenden Eigenschaften, wie die energetische Lage der beiden äußeren Energiebänder, Valenzband und Leitungsband, die Bandlücke (band gap) E_G zwischen den Bändern, die Bandstruktur und die Zustandsdichten. Zur Beschreibung von Transportvorgängen werden quasi-klassische Teilchen eingeführt, die Elektronen im Leitungsband sowie die Löcher im Valenzband. In Analogie zu freien Elektronen können Elektronen im Leitungsband mit einer effektiven Masse beschrieben werden, in der die Wirkung des periodischen Potentials des Kristallgitters berücksichtigt

ist. In Transportmodellen, die auf der Lösung der Boltzmann-Gleichung beruhen, wird der Elektronentransport als eine Folge von Phasen freien Fluges und Stoßprozessen angesehen. Während der Freiflugphasen kann ein Elektron durch ein elektrisches Feld beschleunigt werden, wobei die Gesetze der Newtonschen Mechanik gelten. Ursachen für Stöße können Wechselwirkungen der Elektronen untereinander sein, vor allem aber Wechselwirkungen mit Störungen des idealen Kristallgitters, wie beispielsweise Dotanten, Grenzflächen oder Phononen (das sind Quasiteilchen zur Beschreibung von Gitterschwingungen).

3.1.1. Transportmodelle

Der Transport eines Ensembles von Elektronen lässt sich mit Hilfe der Boltzmann-Transportgleichung (BTG) [67]

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f + \frac{1}{\hbar} \mathbf{F} \cdot \nabla_{\mathbf{k}} f = \left[\frac{\partial f}{\partial t} \right]_{\text{StoB}} \quad (3-1)$$

beschreiben. Dabei ist $f = f(\mathbf{r}, \mathbf{k}, t)$ die Verteilungsfunktion des Ensembles¹. Sie gibt die Wahrscheinlichkeit an, ein Elektron zu einer bestimmten Zeit t an einem bestimmten Ort \mathbf{r} in einem Zustand \mathbf{k} anzutreffen. Der Vektor \mathbf{k} beschreibt den Zustand des Elektrons im Impulsraum. Der Vektor \mathbf{u} ist die Geschwindigkeit des Elektrons im Zustand \mathbf{k} , \hbar ist die reduzierte Planck-Konstante ($\hbar = h / 2\pi$) und $\nabla_{\mathbf{k}}$ der Gradienten-Operator im Impulsraum. Gleichung (3-1) sagt aus, dass eine zeitliche Änderung der Verteilungsfunktion durch Bewegung von Elektronen im Ortsraum infolge von Diffusionsvorgängen, durch Bewegung im Impulsraum infolge äußerer Kräfte \mathbf{F} sowie durch Stöße (innere Kräfte) verursacht werden kann. Da die Beschreibung jedes einzelnen Stoßprozesses viel zu kompliziert wäre, wird der Stoßterm auf der rechten Seite gewöhnlich als Integral über die wahrscheinlichen Übergänge von einem \mathbf{k} -Zustand in einen anderen dargestellt. Der Transport von Löchern kann ebenfalls mit Gleichung (3-1) beschrieben werden.

Die Lösung der BTG ist eine äußerst schwierige Aufgabe. Es handelt sich hierbei um eine so genannte Integro-Differentialgleichung mit sieben unabhängigen Variablen, für die es keine geschlossene Lösung gibt [68]. Durch eine Reihe von zum Teil sehr starken Annahmen lässt sich die BTG jedoch zu Transportmodellen vereinfachen, mit denen

¹ Man bezeichnet ein Ensemble von Elektronen häufig auch als Elektronengas.

sich ein Grossteil der praktisch relevanten Fälle beschreiben lässt. Die verschiedenen Transportmodelle unterscheiden sich im Grad dieser Näherungen.

Das einfachste Transportmodell ist das Drift-Diffusionsmodell (DDM). Es beruht auf der Annahme, dass sich f nicht weit von der Gleichgewichtsverteilung f_0 , das ist die Fermi-Dirac-Verteilung, entfernt. Die Transportgleichung des DDM für Elektronen lautet [68]

$$\mathbf{J}_n = q \cdot n \cdot \mu_n \cdot \mathbf{E} + q \cdot D_n \cdot \nabla n, \quad (3-2)$$

wobei \mathbf{J}_n die Elektronenstromdichte ist, n die Elektronenkonzentration, μ_n die Beweglichkeit, D_n der Diffusionskoeffizient der Elektronen und \mathbf{E} die elektrische Feldstärke. Für D_n gilt die Einsteinbeziehung $D_n = \mu_n \cdot V_T$. Der erste Term auf der rechten Seite von (3-2) beschreibt den Driftstrom, der zweite Term den Diffusionsstrom. Gleichung (3-2) bildet zusammen mit der Transportgleichung für Löcher, den Kontinuitätsgleichungen für Elektronen und Löcher, sowie der Poissonsgleichung

$$\nabla(\varepsilon \cdot \nabla \varphi) = -\rho \quad (3-3)$$

ein Gleichungssystem, das in Bauelementesimulationen numerisch gelöst wird. In Gleichung (3-3) ist ε die Dielektrizitätskonstante, φ das elektrostatische Potential und ρ die Raumladungsdichte. Kennzeichnend für das DDM ist, dass alle Transportvorgänge ausschließlich von lokalen Größen abhängen. In einem homogenen Kristall wird zum Beispiel die Driftgeschwindigkeit am Ort \mathbf{r}

$$\mathbf{v}_n(\mathbf{r}) = \mu_n(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}) \quad (3-4)$$

nur von der lokalen Beweglichkeit und vom lokalen elektrischen Feld bestimmt. Das beschreibt die Situation in Anordnungen sehr gut, in denen sich das elektrische Feld innerhalb von Distanzen von etwa der mittleren freien Weglänge der Elektronen nur schwach ändert. Diese Voraussetzung ist aber zum Beispiel in modernen MOSFETs mit $L_G < 100\text{nm}$ nicht mehr erfüllt. Hier treten nichtlokale Effekte auf, zum Beispiel Velocity Overshoot, bei denen die Driftgeschwindigkeit größer sein kann, als es nach Gleichung (3-4) zu erwarten wäre². Die Ursache dafür liegt in der Tatsache, dass die Driftgeschwindigkeit von der mittleren Energie der Elektronen abhängt, welche sich nicht schlagartig ändern kann.

² Solche Effekte treten zwar auch in längeren Transistoren auf, jedoch sind deren Stärke und Auswirkungen dort wesentlich geringer.

Im Hydrodynamischen Modell (HDM) versucht man solche Effekte zu erfassen, indem man die Energie des Elektronengases berücksichtigt. Die Transportgleichung des HDM für Elektronen lässt sich schreiben als [69]

$$\mathbf{J}_n = q \cdot n \cdot \mu_n \cdot \mathbf{E} + k_B \cdot \nabla(n \cdot \mu_n \cdot T_n), \quad (3-5)$$

wobei die Temperatur des Elektronengases T_n ein Maß für die mittlere Energie der Elektronen ist und sich von der Gittertemperatur T stark unterscheiden kann. Das Gleichungssystem des HDM umfasst neben den Gleichungen für das DDM zusätzlich noch die Energiebalance- und die Energietransportgleichungen für Elektronen und Löcher [70], was einen wesentlich größeren Rechenaufwand zur Folge hat. Im Gegensatz zum DDM wird beim HDM nicht mehr die Nähe zum thermodynamischen Gleichgewicht gefordert. Stattdessen wird angenommen, dass die Form von $f(\mathbf{r}, \mathbf{k}, t)$ auch bei höheren Energien erhalten bleibt [5], [71]-[72]. Inwiefern sich mit dieser Näherung die Verhältnisse beim Elektronentransport in extrem skalierten MOSFETs beschreiben lassen, ist jedoch fraglich [73].

Eine Möglichkeit die BTG sehr genau zu lösen bietet die Monte-Carlo-Methode (MC) [74]. Dabei geht man nicht mehr von makroskopischen Annahmen aus, sondern beschreibt direkt die mikroskopische Bewegung einzelner Ladungsträger im Halbleiterkristall als Folge von freiem Flug und Stößen. Die Auswahl der Stoßmechanismen und deren Wirkung erfolgt zufällig (daher der Name Monte Carlo), jedoch unter der Berücksichtigung ihrer statistischen Wahrscheinlichkeit. Für eine repräsentative Beschreibung des Elektronengases muss diese Berechnung für eine hinreichend große Zahl von Ladungsträgern wiederholt werden. Makroskopische Größen, wie Ladungsträgerdichten oder Stromdichten erhält man durch Summierung über alle Teilchen. Der numerische Aufwand für MC-Simulationen ist enorm. Ein Vergleich der Rechenzeiten von DD-, HD- und MC-Simulationen lässt sich mit einem Verhältnis von 1:6:140 angeben [75]. Daher kommt die MC-Methode für ingenieurmäßige Aufgaben, bei denen etwa das Bauelementedesign in einem weiten Bereich variiert werden muss, nicht in Frage. Stattdessen dienen die Ergebnisse von MC-Simulationen ausgewählter Strukturen gewöhnlich als Referenzgrößen für Berechnungen mit dem DDM oder HDM.

Werden die charakteristischen Bauelementeabmessungen in Transportrichtung kleiner als die De-Broglie-Wellenlänge der Ladungsträger, werden Quanteneffekte wirksam, die sich nicht mehr mit der semiklassischen BTG beschreiben lassen. Dazu gehört zum Beispiel das Source-Drain-Tunneln in extrem skalierten MOSFETs. Hierfür werden

Modelle zur Beschreibung des Quantentransports benötigt. Häufig wird die Methode der Nichtgleichgewichts-Greensfunktionen (nonequilibrium Green's functions, NEGF) verwendet [76]-[80]. Der Vorteil dieser Methode liegt in der rigorosen quantenmechanischen Beschreibung des Elektronentransports. Besonders Bauelemente, in denen der Transport eindimensional betrachtet werden kann, wie etwa Resonanz-Tunneldioden [81] aber auch SG SOI bzw. MG MOSFETs mit extrem dünnen Kanalschichten ($<5\text{nm}$) [79] lassen sich damit gut beschreiben. Wird eine vollständig zweidimensionale Behandlung des Transports notwendig, wie z.B. bei Bulk-MOSFETs oder SOI MOSFETs mit dickeren Schichten, erhöht sich der numerische Aufwand beträchtlich [77]-[79].

Die NEGF-Methode ist dann besonders vorteilhaft, wenn Streuprozesse vernachlässigt werden können. In Silizium MOSFETs ist Streuung aber bis zu Kanallängen von mindestens 2nm relevant [82]. Eine rigorose Berücksichtigung von Streuprozessen ist mit dem NEGF-Formalismus zwar prinzipiell möglich [83]-[84], aufgrund des enormen Rechenaufwandes beschränkt man sich aber in der Praxis entweder auf einzelne ausgewählte Streumechanismen [85] oder auf eine phänomenologische Beschreibung der Wirkung von Streuung mit Hilfe sogenannter Büttiker-Probes [79], [83], [86].

Alternativ zum NEGF-Formalismus lässt sich Quantentransport auch mit Hilfe der Wigner-Funktion behandeln. Sie erlaubt die Beschreibung von Streuung in einer Form vergleichbar der BTG [87]. Die resultierende Transportgleichung lässt sich daher ebenfalls mit einer MC-Methode lösen [88].

Bei den in dieser Arbeit untersuchten MOSFET-Strukturen sind Effekte des Quantentransports vernachlässigbar gering. Für die Simulationen wurden deshalb die semiklassischen Transportmodelle DDM, HDM und MC verwendet.

3.1.2. Modellierung von Quantum Confinement

Unabhängig von den Transportmodellen gibt es verschiedene Möglichkeiten, die Auswirkung des Quantum Confinement im Kanal, senkrecht zum Stromfluss, zu berücksichtigen. Das genaueste Verfahren hierfür ist die gekoppelte selbstkonsistente Lösung der Schrödinger- und Poissongleichungen in den Richtungen eingeschränkter Bewegungsfreiheit. Als Lösung erhält man die Subbandenergien, die dazugehörigen Wellenfunktionen sowie die damit verbundene räumliche Verteilung der Elektronen. Da sich im Allgemeinen die Form des Potentialgrabens, in dem sich die Elektronen von

Source nach Drain bewegen, entlang des Weges ändert, muss diese Rechnung in einer ausreichenden Anzahl senkrechter Schnittebenen zwischen Source und Drain durchgeführt werden. Das bedeutet einen erheblichen zusätzlichen Rechenaufwand in den Bauelementesimulationen. Daher wird häufig versucht, mit Näherungslösungen zu arbeiten.

Für hoch dotierte Bulk-MOSFETs liefert das van-Dort Modell recht gute Ergebnisse [89]-[90]. Die Grundannahme dieses Modells ist ein dreieckiges Querschnittsprofil des Kanalpotentials, dessen Öffnungswinkel durch das senkrechte elektrische Feld bestimmt ist. Auf DG MOSFETs oder FD SOI MOSFETs lässt sich dieses Modell daher nicht anwenden (vgl. Abb. 2.13).

Eine Methode für beliebige Potentialverläufe bietet das Density-Gradient-Modell. Die Anwendbarkeit dieses Modells auf DG und SG SOI MOSFETs wurde bereits erfolgreich demonstriert [90]. Allerdings hat das Density-Gradient-Modell nicht in jedem Fall Vorteile bezüglich Rechenzeit und numerischer Stabilität im Vergleich zu selbstkonsistenten Schrödinger-Poisson (SP) Simulationen [90].

Eine einfache aber effektive Möglichkeit Quanteneffekte zu berücksichtigen bieten analytische Näherungslösungen, mit denen klassisch simulierte Kennlinien korrigiert werden können. Derartige Modelle existieren sowohl für eindimensionales Quantum Confinement in SG oder DG MOSFETs [65] als auch für zweidimensionales Confinement in MG MOSFETs mit rechteckigem Querschnitt [66]. Im Rahmen dieser Arbeit werden Quanteneffekte auf diese Art und Weise berücksichtigt. Die Genauigkeit der analytischen Modelle wird in Abschnitt 3.4. durch einen Vergleich mit selbstkonsistenten Schrödinger-Poisson Simulationen überprüft.

Darüber hinaus wird in Abschnitt 3.4.2. die Auswirkung des Quantum Confinement auf die Gatekapazität sowie in Abschnitt 3.4.3. auf die Elektronenbeweglichkeit in Tri-Gate Nanowire MOSFETs mit Hilfe von zweidimensionalen SP-Simulationen untersucht.

3.2. Simulation des Ladungstransports in Nanometer-MOSFETs

3.2.1. Vergleich der Transportmodelle

Für die Untersuchungen zum Design extrem skaliert MOSFETs im Rahmen dieser Arbeit wurde ein Simulationsmodell benötigt, welches numerisch stabil ist und mit geringem Rechenaufwand das Gleichstromverhalten hinreichend genau beschreibt. Dafür kamen im Prinzip nur das DDM und das HDM in Frage. Um die Eignung beider Modelle zu testen, wurden Vergleichssimulationen an ausgewählten zweidimensionalen MOSFET-Strukturen durchgeführt. Als Referenz dienten MC-Simulationen dieser Strukturen. Die wichtigsten Ergebnisse dieses Vergleichs wurden bereits veröffentlicht [91]-[92] und werden im Folgenden noch einmal vorgestellt und diskutiert.

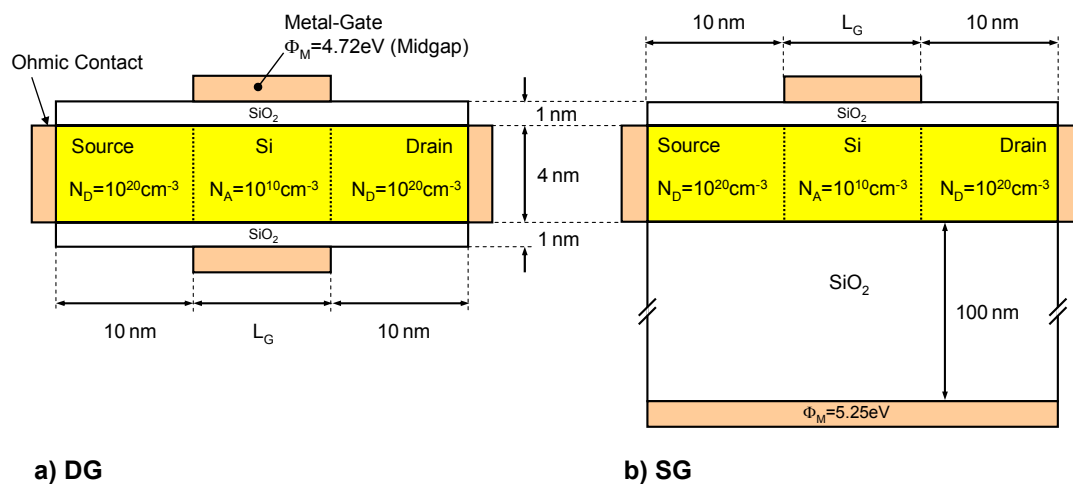


Abbildung 3.1 Querschnitte der simulierten MOSFET-Strukturen für den Vergleich der Transportmodelle. (a) Double-Gate MOSFET. (b) Single-Gate SOI MOSFET.

Die Simulationsstrukturen sind in Abbildung 3.1 skizziert. Es handelt sich um n-Kanal SG und DG SOI MOSFETs mit undotiertem Kanal und Metall-Gates. Es wurde eine Gateaustrittsarbeit von 4.72 eV angenommen, also die von Midgap-Material. Die Gatelängen dieser Transistoren wurden in einem weiten Bereich zwischen 5 und 100 nm variiert, während alle übrigen Strukturparameter konstant blieben. Das Gateoxid und die Siliziumschicht sind extrem dünn mit Schichtdicken von $t_{ox} = 1 \text{ nm}$ und $t_{Si} = 4 \text{ nm}$. Dadurch wird ein Abschalten der MOSFETs auch bei sehr geringer Gatelänge

ermöglicht. Die Dotierung von Source und Drain beträgt 10^{20}cm^{-3} und die Übergänge zum undotierten Kanal sind abrupt. Durch dieses idealisierte Design sind die Serienwiderstände der Source- und Draingebiete minimal, so dass die Unterschiede zwischen den Transportmodellen nicht durch parasitäre Effekte überdeckt werden.

Die MC-Simulationen wurden mit dem Ensemble-Monte-Carlo-Simulator DAMOCLES [6], [93] durchgeführt. Dabei wurde ein analytisches nichtparabolisches Modell für die Bandstruktur von Silizium verwendet und die Streuung an Phononen, geladenen und neutralen Störstellen sowie an Grenzflächen berücksichtigt. Quantisierung senkrecht zum Stromfluss wurde nicht betrachtet, da es in dieser Studie in erster Linie um die Beschreibung des Ladungstransports ging.

Für die DD- und HD-Simulationen wurde der kommerzielle Bauelementesimulator ATLAS [5] verwendet. Zur Beschreibung der Niedrigfeldbeweglichkeit unter Berücksichtigung des senkrechten Feldes wurde das Lombardi-Modell [94] eingeschaltet. Innerhalb des DDM wird das Sättigungsverhalten der Driftgeschwindigkeit bei hohen Feldstärken in Transportrichtung mit der $v(E)$ -Charakteristik nach Caughey und Thomas [95]

$$v(E) = \mu_{\text{eff}} E \left(\frac{1}{1 + (\mu_{\text{eff}} E / v_{\text{sat}})^\beta} \right)^{1/\beta} \quad (3-6)$$

beschrieben. Dabei ist v die Geschwindigkeit der Ladungsträger, E die elektrische Feldstärke in Transportrichtung und μ_{eff} die effektive Niedrigfeldbeweglichkeit aus dem Lombardi-Modell. Der Parameter β bestimmt den Übergang zwischen Niedrigfeldtransport mit der Geschwindigkeit $v \approx \mu_{\text{eff}} E$ und Hochfeldtransport mit der Sättigungsgeschwindigkeit v_{sat} . Die Standardparameter in ATLAS sind $v_{\text{sat}} = 10^7\text{cm/s}$ und $\beta = 2$ für Elektronen.

Für das HDM wird noch ein weiterer wichtiger Parameter benötigt, die Energierelaxationszeit τ_w . Sie ist ein Maß für die mittlere Zeit, in der nach dem Abschalten äußerer Felder das aufgeheizte Elektronengas in das thermodynamische Gleichgewicht zurückkehrt. Der Standardwert für τ_w in ATLAS ist 0.4ps.

Für einen konsistenten Vergleich der Transportmodelle ist es wichtig, dass sie in Fällen geringer Abweichungen vom thermodynamischen Gleichgewicht, also bei kleinen Feldstärken in Transportrichtung, die gleichen Ergebnisse liefern. Deshalb wurden die Parameter des Lombardi-Modells so justiert, dass das DDM die gleichen Transferkennlinien für SG und DG MOSFETs mit $L_G = 100\text{nm}$ bei $V_{\text{DS}} = 50\text{mV}$ liefert

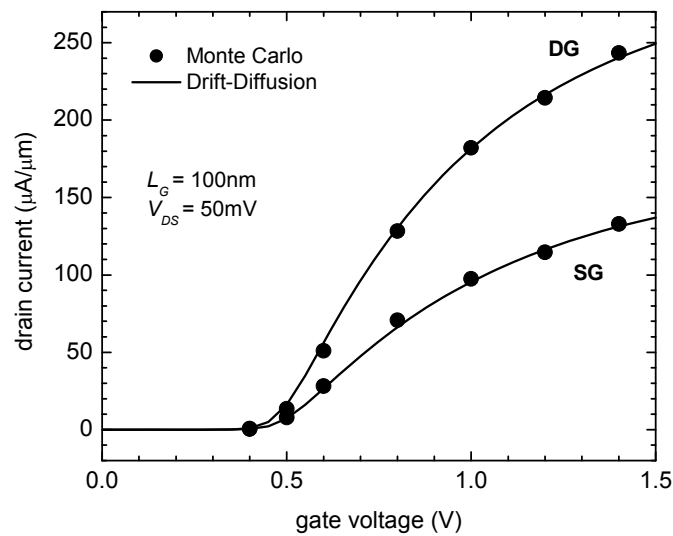


Abbildung 3.2 Simulierte Transferkennlinien für SG und DG MOSFETs mit $L_G = 100\text{nm}$ bei geringer Drainspannung ($V_{DS} = 50\text{mV}$). Zu sehen sind DD- und MC-Ergebnisse nach der Feinjustierung des Lombardi-Modells für die Niedrigfeldbeweglichkeit in den DD-Simulationen.

wie die MC-Simulationen (Abb. 3.2). Alle weiteren Simulationen wurden dann für eine Drainspannung von 1V durchgeführt. In allen Simulationsmodellen (DDM, HDM, MC) wurde Fermi-Dirac-Statistik verwendet.

Zunächst wurden für alle MOSFET-Strukturen die Transferkennlinien mit dem DDM und dem HDM unter Verwendung der ATLAS-Standardparameter (mit Ausnahme des justierten Lombardi-Modells) sowie mit MC simuliert. Als ein Beispiel sind in Abbildung 3.3 die simulierten Transferkennlinien für DG MOSFETs mit Gatelängen von 10 und 100nm zu sehen. Für den 100nm-Transistor liefern alle Modelle vergleichbare Drainströme, sowohl im Subthresholdbereich [Abb. 3.3(a)] als auch im On-Zustand [Abb. 3.3(b)]. Vom HDM wird der On-Strom, $I_{on} = I_D(V_{DS} = V_{GS} = 1\text{V})$, um etwa 15% überschätzt. Für den 10nm-MOSFET liefert das HDM deutlich zu große Ströme über den gesamten Bereich der Gatespannung. Hier ist der vom HDM vorhergesagte On-Strom um 54% größer als der aus MC-Simulationen. Auch die Subthresholdströme aus dem HDM sind um etwa 43% zu groß. In diesen Trend passt eine kürzlich veröffentlichte Studie [75], in der die Drainströme eines 60nm-Bulk-MOSFET vom HDM um 30% überschätzt wurden. Diese Überbestimmung des Drainstromes durch das HDM hängt mit dessen Unfähigkeit zusammen, die ballistische Injektion von Ladungsträgern in den Kanal qualitativ korrekt zu beschreiben [96]-[97].

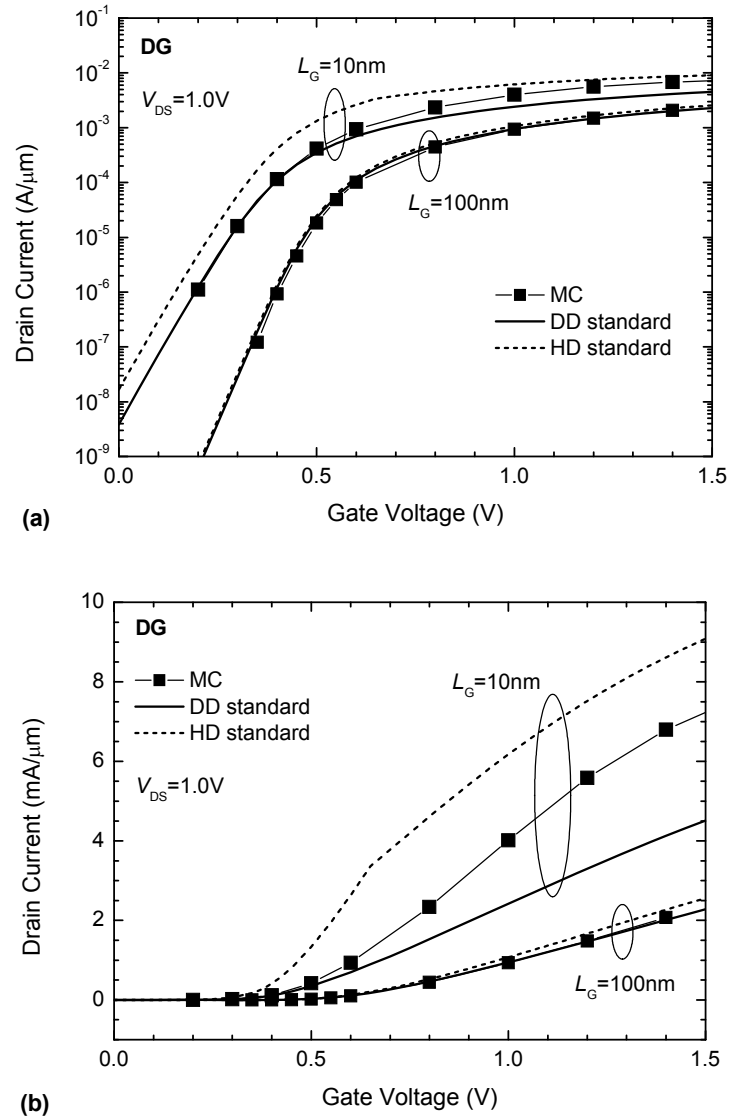


Abbildung 3.3 Transferkennlinien von DG MOSFETs mit $L_G = 10\text{nm}$ und 100nm aus DD-, HD- und MC-Simulationen. (a) Logarithmischer Maßstab. (b) Linearer Maßstab.

Wie in Abbildung 3.3 zu sehen ist, liefert das DDM für den 100nm -DG MOSFET nahezu die gleiche Transferkennlinie wie die MC-Simulationen. Für den 10nm -Transistor wird der On-Strom vom DDM jedoch um 40% unterschätzt. Dieser Unterschied zwischen DDM und MC verdeutlicht den Einfluss von nichtstationärem Transport und ballistischen Effekten bei derart kurzen Kanälen. Im Subthresholdbereich stimmen DD- und MC-Ergebnisse wiederum sehr gut überein. Die Simulationen der SG MOSFETs ergaben ähnliche Trends bezüglich des Vergleiches von DDM und HDM mit MC.

Gründe dafür, dass das DDM auch bei MOSFETs mit Gatelängen deutlich unter 100nm zum Teil gute Ergebnisse liefern kann, wurden bereits von Banoo, Lundstrom und Ren [96], [98] diskutiert. Demnach wird die Höhe des Drainstromes durch die ballistische Injektion von Elektronen über die Potentialbarriere am Source-Ende des Kanals und deren Rückstreuung bestimmt. Nach der von Lundstrom und Mitarbeitern entwickelten Streuungstheorie [99]-[103] lässt sich die mittlere Elektronengeschwindigkeit auf dem Maximum der Potentialbarriere, also am Source-Ende des Kanals, mit Gleichung (3-7) beschreiben [98].

$$v(0) = \mu_{\text{eff}} E \frac{1}{1 + (\mu_{\text{eff}} E / v_T)} \quad (3-7)$$

Hierbei ist v_T die thermische Geschwindigkeit der Source-Elektronen in Transportrichtung und E das parallele elektrische Feld am Beginn des Kanals. Die Ähnlichkeit von (3-7) zu Gleichung (3-6), welche die Geschwindigkeitssättigung bei hohen parallelen Feldern innerhalb des DDM beschreibt, ist verblüffend. Man erhält Gleichung (3-7) aus (3-6) leicht, indem man $\beta = 1$ setzt und v_{sat} durch v_T ersetzt. Die thermische Geschwindigkeit v_T hängt vom Grad der Entartung des Elektronengases ab, das heißt von der Elektronendichte am Beginn des Kanals. Im nichtentarteten Grenzfall, also bei geringer Elektronendichte, liegt der Wert von v_T etwas über 10^7 cm/s [101], [103], also nahe der Sättigungsgeschwindigkeit v_{sat} . Die Ähnlichkeit von v_{sat} zu v_T erklärt die gute Übereinstimmung der DD- und MC-Kennlinien für den 10nm-MOSFET im Subthresholdbereich [Abb. 3.3(a)], da hier die Elektronendichte im Kanal gering ist. Bei hohen Konzentrationen ($n_S > 10^{12} \text{ cm}^{-2}$) wächst v_T exponentiell mit der Elektronendichte und kann deutlich größer sein als v_{sat} [101]. Daher ist der On-Strom für den 10nm-Transistor [Abb. 3.3(b)] aus dem DDM auch wesentlich kleiner als der aus MC-Simulationen.

3.2.2. Anpassung des DDM und des HDM

Um die Genauigkeit des DDM bei der Simulation von On-Strömen extrem skalierten MOSFETs zu erhöhen ist es nahe liegend, die Parameter β und v_{sat} in Gleichung (3-6) zu justieren. Setzt man $\beta = 1$, lässt sich für jede Gatelänge ein v_{sat} bestimmen, mit dem das DDM den gleichen On-Strom liefert wie die MC-Methode. Abbildung 3.4 zeigt die so ermittelten v_{sat} für die untersuchten SG und DG MOSFETs als Funktion der Gatelänge. Darin ist zu sehen, dass die optimale Sättigungsgeschwindigkeit für DG

MOSFETs etwas größer ist als für SG MOSFETs. Für alle Gatelängen sind die Unterschiede jedoch relativ gering. Die für beide MOSFET-Typen gemittelte Abhängigkeit $v_{\text{sat}}(L_G)$ lässt sich mit der Funktion

$$v_{\text{sat}}(L_G) = a + \frac{b}{L_G^c} \quad (3-8)$$

beschreiben. Mit den Fitparametern $a = 2.0$, $b = 19.2$ und $c = 1.43$ liefert Gleichung (3-8) für jedes L_G in nm ein mittleres v_{sat} in cm/s. Dieses gatelängenabhängige v_{sat} kann nun in (3-6) verwendet werden, um mit dem DDM extrem skalierte MOSFETs zu simulieren.

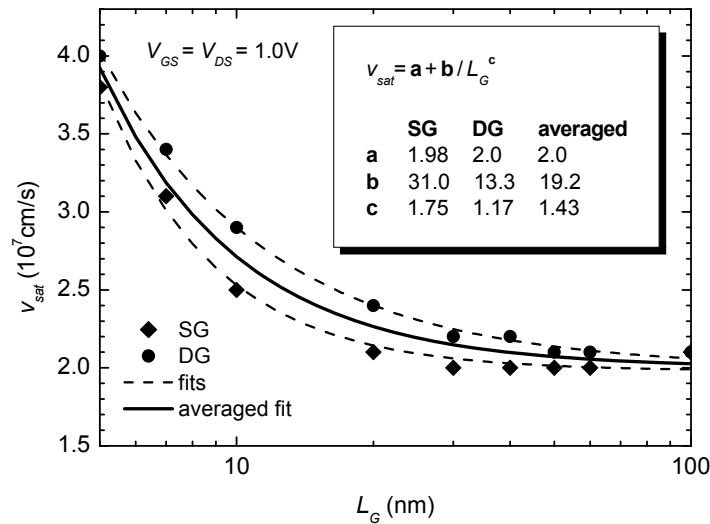


Abbildung 3.4 Optimale Sättigungsgeschwindigkeit für SG und DG MOSFETs (Symbole) als Funktion der Gatelänge. Außerdem gezeigt sind ein mittlerer Fit mit Gleichung (3-8) (durchgezogene Linie), sowie spezielle Fits für SG und für DG MOSFETs (gestrichelte Linien).

Abbildung 3.5 zeigt einen Vergleich der mit dem DDM und mit MC simulierten On-Ströme für SG und DG MOSFETs als Funktion der Gatelänge. In den DD-Simulationen wurden drei verschiedene Parametersätze für Gleichung (3-6) verwendet: die ATLAS-Standardparameter $\beta = 2$ und $v_{\text{sat}} = 10^7$ cm/s, die von Bude vorgeschlagenen Parameter $\beta = 1$ und $v_{\text{sat}} = 2.2 \times 10^7$ cm/s [104] sowie die optimierten Parameter $\beta = 1$ und $v_{\text{sat}}(L_G)$ nach Gleichung (3-8). Alle Transportmodelle sagen eine qualitativ ähnliche Gatelängenabhängigkeit des On-Stroms voraus. Für SG oder DG MOSFETs mit $L_G > 40$ nm sind die On-Ströme aus dem DDM mit ATLAS-Standardparametern nur geringfügig (weniger als 10%) kleiner als die MC-Ströme. Für kürzere Transistoren liefert das DDM allerdings deutlich kleinere Ströme als MC. Bei $L_G = 20$ nm berechnet

das DDM nur noch 75% des MC-Stromes, bei $L_G = 10\text{nm}$ sind es gerade noch 60%. Damit wird der zunehmende Einfluss ballistischer Effekte offensichtlich. Bude's Parametersatz [104] für das DDM liefert gute Übereinstimmung mit MC-Ergebnissen bis zu $L_G = 20\text{nm}$. Bei Verwendung von $v_{\text{sat}}(L_G)$ nach Gleichung (3-8) werden die MC-Ströme vom DDM über den gesamten Gatelängenbereich von 5 bis 100nm sehr gut wiedergegeben. Vergleichssimulationen haben gezeigt, dass sich diese Aussage nicht wesentlich ändert, wenn eine äquivalente Oxiddicke von 2nm angenommen wird oder die Drainspannung um $\pm 0.2\text{V}$ variiert.

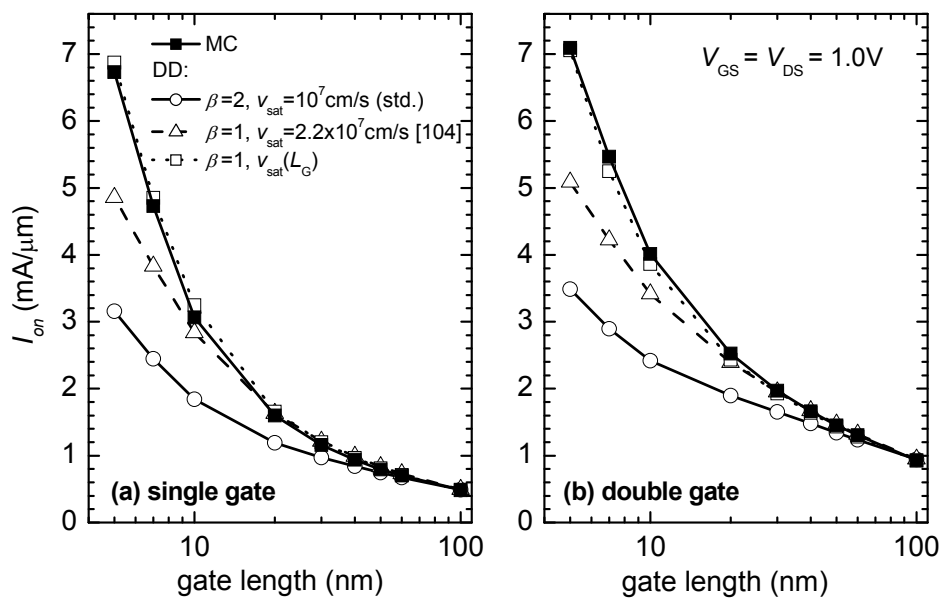


Abbildung 3.5 Mit dem DDM simulierte On-Ströme von (a) SG und (b) DG MOSFETs als Funktion der Gatelänge im Vergleich zu MC-Ergebnissen. Für das DDM wurden drei verschiedene Parametersätze verwendet: die ATLAS-Standardparameter, Bude's Parameter [104], sowie die hier vorgestellten Parameter mit $v_{\text{sat}}(L_G)$ nach Gleichung (3-8).

Es muss darauf hingewiesen werden, dass die Modifikation des DDM nicht mit dem Ziel gemacht wurde, die Physik des Ladungstransports in extrem skalierten MOSFETs korrekt zu beschreiben. Es ging dabei ausschließlich um die Nachbildung realistischer Ströme. Betrachtet man jedoch die simulierten Elektronengeschwindigkeiten im Kanal, so stellt man eine erstaunlich gute Übereinstimmung zu den MC-Ergebnissen fest. In Abbildung 3.6 sind die mit den drei Varianten des DDM sowie mit MC simulierten Geschwindigkeitsprofile entlang des Kanals eines 10nm-DG MOSFETs zu sehen. In dieser Darstellung ist die Elektronengeschwindigkeit der jeweilige Mittelwert aller Elektronen am Ort x . Das mit $v_{\text{sat}}(L_G)$ simulierte Geschwindigkeitsprofil gibt das MC-

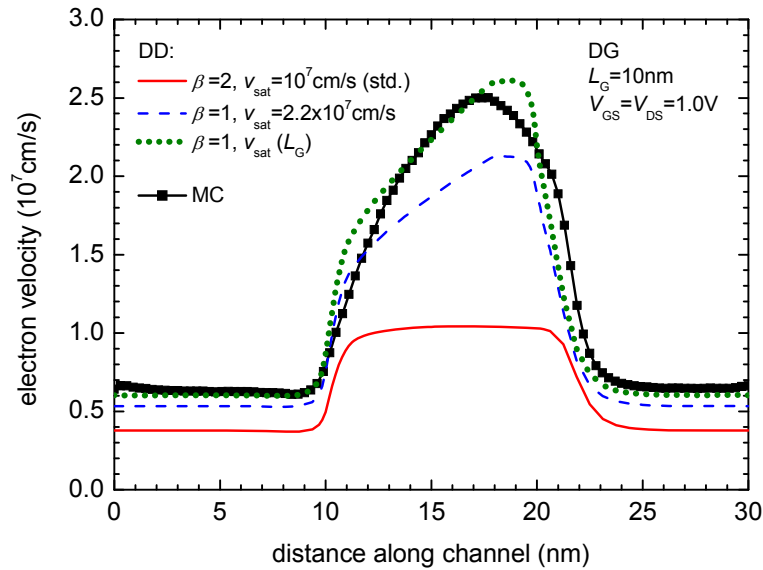


Abbildung 3.6 Mittlere Elektronengeschwindigkeit entlang des Kanals eines DG MOSFETs mit $L_G = 10\text{nm}$. Die Geschwindigkeitsprofile wurden mit den drei Varianten des DDM sowie mit MC simuliert.

Profil sehr gut wieder. Das ist nicht der Fall mit den beiden anderen Parametersätzen für das DDM.

Die Genauigkeit des HDM lässt sich über eine Anpassung der Energierelaxationszeit τ_w verbessern. Um die Überbestimmung der On-Ströme durch das HDM zu verringern, musste der ATLAS-Standardwert von $\tau_w = 0.4\text{ps}$ deutlich verkleinert werden. Es hat sich herausgestellt, dass sich für jeden MOSFET-Typ ein konstantes τ_w finden lässt, mit dem das HDM die MC-On-Ströme für Gatelängen zwischen 10 und 100nm sehr gut wiedergibt (Abb. 3.7). Die optimalen Werte sind 0.09ps für SG MOSFETs und 0.13ps für DG MOSFETs. Diese Werte sind dem τ_w aus [105] sehr ähnlich. Trotz der guten Übereinsimmung der mit dem justierten HDM und mit MC simulierten On-Ströme sind die Unterschiede in den berechneten Elektronengeschwindigkeiten weiterhin recht groß. Wie in Abbildung 3.8 zu sehen ist, liefert auch das justierte HDM im Vergleich zu MC deutlich zu hohe Peakgeschwindigkeiten. Wie von Lundstrom u. a. [96], [100]-[101] gezeigt wurde, ist für die Höhe des Drainstromes nicht die Peakgeschwindigkeit entscheidend sondern die Injektions-geschwindigkeit am Beginn des Kanals. In Abbildung 3.8 wird deutlich, dass die Elektronengeschwindigkeiten in der Nähe des Source-Kanal-Übergangs vom justierten HDM und MC sehr ähnlich berechnet werden, was die gute Übereinstimmung der On-Ströme erklärt.

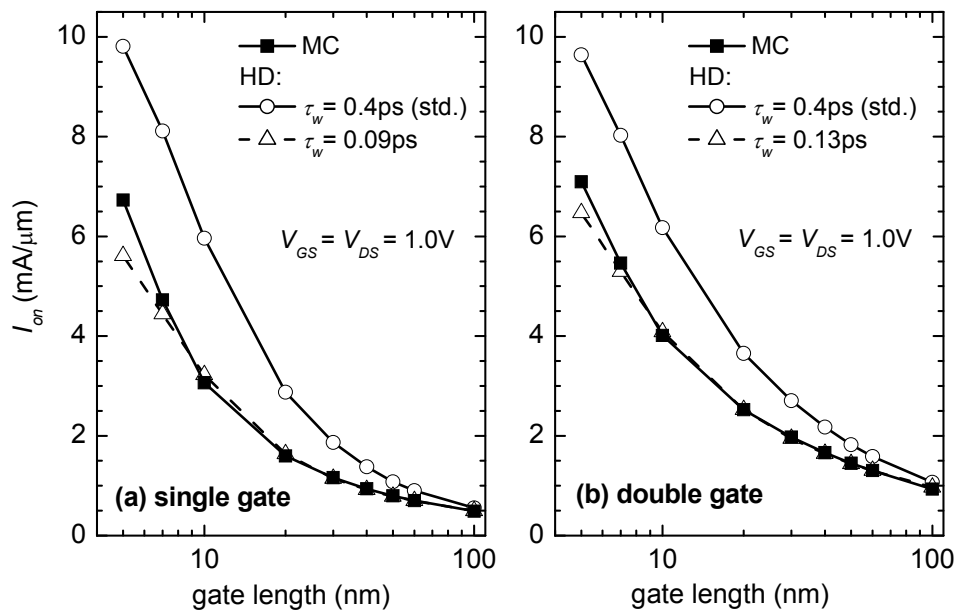


Abbildung 3.7 Mit dem HDM simulierte On-Ströme von (a) SG und (b) DG MOSFETs als Funktion der Gatelänge im Vergleich zu MC-Ergebnissen. Für die HD-Simulationen wurden verschiedene Energielaxationszeiten benutzt: der ATLAS-Standardwert $\tau_w = 0.4ps$ sowie optimierte Werte für SG MOSFETs ($\tau_w = 0.09ps$) und DG MOSFETs ($\tau_w = 0.13ps$).

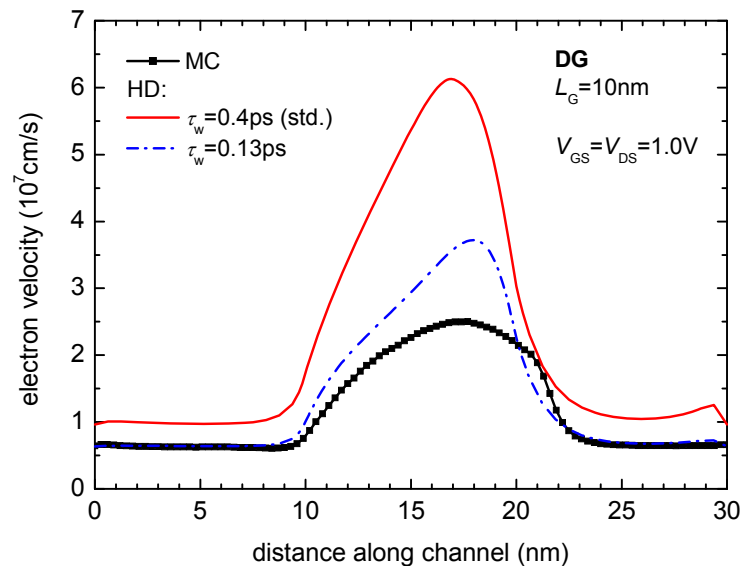


Abbildung 3.8 Mittlere Elektronengeschwindigkeit entlang des Kanals eines DG MOSFETs mit $L_G = 10nm$ aus Simulationen mit dem HDM (standard und justiertes τ_w) sowie mit MC.

Für die Auswahl eines Simulationsmodells zur Beschreibung extrem skaliert MOSFETs ist die korrekte Wiedergabe des On-Stromes wichtig, aber nicht das einzige Kriterium. Ebenfalls von Bedeutung ist eine hinreichend genaue Beschreibung des Subthresholdverhaltens sowie des Übergangsbereiches zwischen On- und Off-Zustand. Abbildung 3.9 zeigt einen Vergleich der mit den modifizierten DD- und HD-Modellen sowie mit MC simulierten Transferkennlinien eines 10nm-DG MOSFETs. Es ist zu sehen, dass die Kennlinien des DDM und des HDM abgesehen vom On-Strom mehr oder weniger stark von der MC-Kennlinie abweichen. Die DD-Kennlinie liegt jedoch im Vergleich zur HD-Kurve sehr dicht an den MC-Daten über den gesamten Bereich der Gatespannung. Die Ungenauigkeiten des modifizierten DDM sind im Wesentlichen ein etwas zu hoher Off-Strom sowie die geringere Steilheit im Bereich von I_{on} . Die Ausgangskennlinien (Abb. 3.10) werden vom modifizierten DDM sehr gut wiedergegeben, während das justierte HDM auch hier recht starke Abweichungen zu den MC-Daten liefert. Der Vergleich spricht also eindeutig für das modifizierte DDM. Die Ungenauigkeiten, die damit bei der Simulation der Gleichstromkennlinien extrem skaliert MOSFETs gemacht werden, sind für die Design-Untersuchungen im Rahmen dieser Arbeit vertretbar. Für diese Untersuchungen wurde deshalb das modifizierte DDM verwendet.

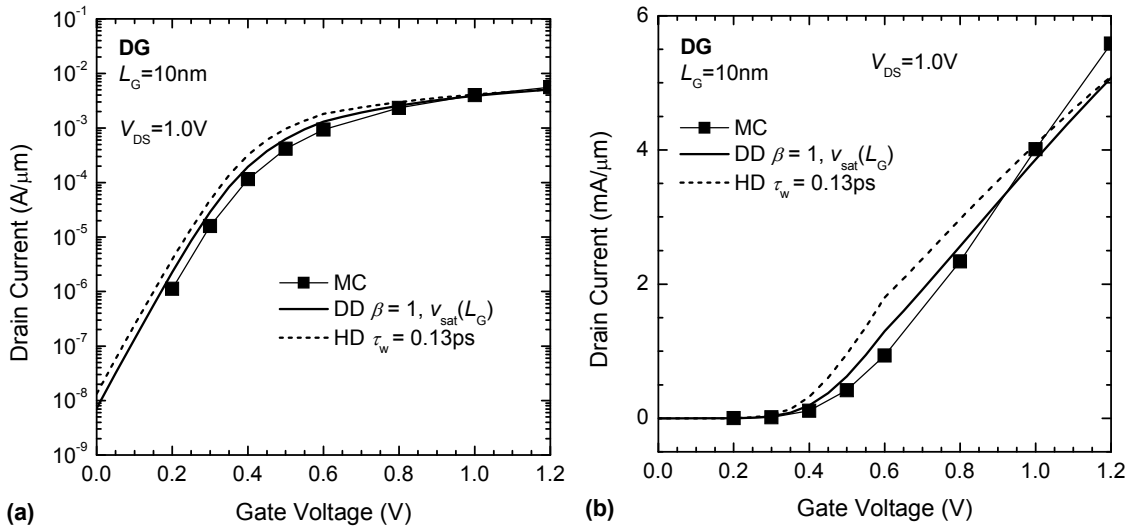


Abbildung 3.9 Mit dem modifizierten DDM, dem modifizierten HDM sowie mit MC simulierte Transferkennlinien eines DG MOSFETs mit $L_G = 10\text{nm}$ im (a) logarithischen und (b) linearen Maßstab.

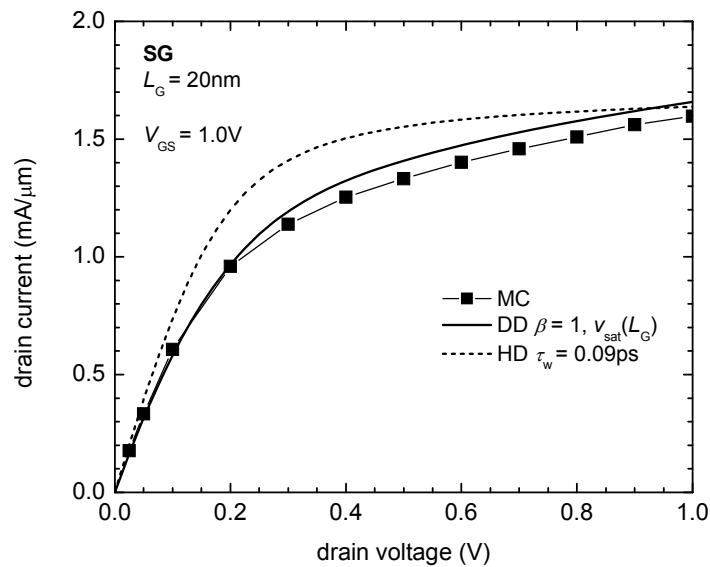


Abbildung 3.10 Mit dem modifizierten DDM, dem modifizierten HDM sowie mit MC simulierte Ausgangskennlinien eines SG MOSFETs mit $L_G = 20\text{nm}$.

3.3. Simulation der Hochfrequenzeigenschaften

Einer der Schwerpunkte dieser Arbeit liegt in der Untersuchung der Hochfrequenzeigenschaften extrem skalierten MOSFETs. Die dafür infrage kommenden Methoden der Kleinsignalanalyse und Simulationsmodelle werden in den folgenden Abschnitten vorgestellt und diskutiert.

3.3.1. Methoden der Kleinsignalanalyse

Grundsätzlich hat die Kleinsignalanalyse eines Bauelements mit mehreren Anschlüssen zum Ziel, dessen Admittanzmatrix³ zu bestimmen. Ein MOSFET ist ein Bauelement mit drei oder auch vier Anschlüssen, je nachdem ob ein Bulk-Anschluss vorhanden ist oder nicht. Die Elemente der Admittanzmatrix lauten in komplexer Schreibweise [106]

$$\tilde{Y}_{ij} = \left. \frac{\tilde{I}_i}{\tilde{V}_j} \right|_{\tilde{V}_k=0, k \neq j}. \quad (3-9)$$

³ Die Admittanz ist der komplexe Leitwert.

Die \tilde{Y}_{ij} beschreiben die Stromänderung am Anschluss i bei einer Spannungsänderung am Anschluss j , während an allen anderen Anschlüssen die angelegte Spannung konstant ist. Die \tilde{Y}_{ij} sind grundsätzlich abhängig vom Gleichstromarbeitspunkt und von der Frequenz. In der Admittanzmatrix sind auch die Y-Parameter enthalten, wie sie mit Gleichung (2-10) definiert wurden. Durch die Zerlegung der \tilde{Y}_{ij} in Real- und Imaginärteil entsprechend

$$\tilde{Y}_{ij} = G_{ij} + j\omega C_{ij} \quad (3-10)$$

sind die Elemente der Leitwert- und der Kapazitätsmatrix definiert. Verwendet man für die Indizes i und j die Anfangsbuchstaben der jeweiligen MOSFET-Anschlüsse ($i, j = G, D, S$), so entspricht in dieser Schreibweise z.B. der Leitwert G_{DG} genau der Steilheit g_m nach Gleichung (2-4). Auch C_{GS} und C_{GD} entsprechen genau den mit (2-18) und (2-19) definierten Kapazitäten.

Zur Bestimmung der Admittanzmatrix mit Hilfe numerischer Simulationen kommen im Wesentlichen drei Methoden infrage, die nachfolgend kurz vorgestellt werden.

A. Fourieranalyse

Die Fourieranalyse ist prinzipiell immer anwendbar, Voraussetzung ist jedoch ein Code zur Simulation des transienten Verhaltens. Die Methode funktioniert nach folgendem Schema: Am Kontakt j wird ein kleiner Spannungssprung angelegt und die Sprungantwort, d.h. der zeitliche Verlauf der Ströme an den Anschlüssen i , wird simuliert. Eine anschließende Fourieranalyse liefert die \tilde{Y}_{ij} als Funktion der Frequenz. Dieses Verfahren ist sehr rechenintensiv. Darüber hinaus sind die Wahl der Amplitude ΔV_j des Spannungssprungs sowie die optimale Wahl des Zeitschritts zur Berechnung der Sprungantwort entscheidend für die Genauigkeit dieser Methode [106]. So ist z.B. die Generation von Harmonischen infolge des endlichen ΔV_j nicht ausgeschlossen.

B. Sinusoidal Steady State Analysis (S^3A)

Die S^3A -Methode ist das genaueste Verfahren zur Bestimmung der Kleinsignalparameter und wird daher auch als Referenzverfahren verwendet [106]. Das Prinzip ist folgendes: Im Gleichstromarbeitspunkt wird der MOSFET nacheinander an jedem Kontakt mit einer sinusförmigen Wechselspannung angeregt und die

verursachten Wechselströme an den einzelnen Kontakten werden berechnet. Daraus wird entsprechend Gleichung (3-9) die Admittanzmatrix des MOSFET bestimmt. Für die Berechnung werden nur die linearen Terme der Taylorentwicklung der Halbleitergleichungen verwendet, mit anderen Worten die ersten Ableitungen. Dadurch ist die Generation von Harmonischen infolge nichtlinearer Verzerrungen ausgeschlossen und die Methode ist unabhängig von der Signalamplitude. Die S^3A -Methode liefert daher exakt die Kleinsignalparameter für die betrachtete Frequenz. Bei Simulationen mit dem DDM oder dem HDM ist eine Kleinsignalanalyse mit S^3A ohne weiteres möglich. Das gilt jedoch nicht für MC-Simulationen. Es gibt zwar bereits Ansätze, die S^3A -Methode auch mit Monte-Carlo zu vereinbaren [74], diese beziehen sich jedoch bisher nur auf homogenes (Bulk-) Material und nicht auf Bauelemente.

C. Charge Partitioning (CP)

Mit Hilfe der CP-Methode lassen sich die Kleinsignalparameter mit sehr geringem Rechenaufwand bestimmen. Man benötigt dafür lediglich einen Code zur Berechnung von Gleichstromarbeitspunkten. Allerdings ist die CP-Methode nicht in jedem Fall anwendbar. Sie kann aber für bestimmte Bauelemente gute Ergebnisse liefern, z.B. für MOSFETs [106]-[107] oder für Bipolartaransistoren [108]-[109]. Problematisch ist im Allgemeinen die korrekte Bestimmung der Kapazitätsmatrix [106].

Die CP-Methode funktioniert nach folgendem Prinzip: Am Kontakt j wird die Spannung um einen kleinen Schritt ΔV_j geändert. Die Änderung der Ströme an den Anschlüssen i sowie die Änderung der Ladung im Bauelement werden ermittelt. Die Ladung muss dann den einzelnen Anschlüssen zugeordnet werden. Diese Zuordnung ist in bestimmten Fällen eindeutig möglich, im Allgemeinen jedoch nicht. Die einzelnen Leitwerte ergeben sich dann aus

$$G_{ij} = \frac{\Delta I_i}{\Delta V_j}, \quad (3-11)$$

die Kapazitäten aus

$$C_{ij} = \frac{\Delta Q_i}{\Delta V_j}. \quad (3-12)$$

Beim MOSFET lassen sich so z.B. die Kapazitäten C_{GS} und C_{GD} eindeutig bestimmen.

3.3.2. Vergleich der Transportmodelle

In Abschnitt 3.2.2 wurde gezeigt, dass es durch geeignete Modifikationen der Modellparameter möglich ist, auch mit dem DDM oder dem HDM die On-Ströme extrem skaliert MOSFETs quantitativ korrekt zu simulieren. Sind die beiden Modelle aber auch für die Simulation des Kleinsignalverhaltens solcher Transistoren geeignet? Zur Klärung dieser Frage wurde ein erneuter Vergleich von DD-, HD- und MC-Simulationen durchgeführt. Vergleichsgegenstand war diesmal die Transitfrequenz f_T eines DG MOSFET entsprechend Abbildung 3.1 mit $L_G = 10\text{nm}$.

Da das Referenzverfahren MC zur Simulation des Ladungstransports nicht mit dem Referenzverfahren S³A zur Simulation des Kleinsignalverhaltens kompatibel ist, wurden außer S³A auch drei Varianten von CP zur Bestimmung von f_T verwendet. Die S³A-Methode kam also nur bei den DD- und HD-Simulationen zur Anwendung. Dabei wurden die Kleinsignalparameter für eine Frequenz von 100GHz simuliert und anschließend f_T nach Gleichung (2-15) bestimmt.

Die CP-Methode kann direkt zur Bestimmung von f_T verwendet werden, wenn die Serienwiderstände des untersuchten MOSFET vernachlässigbar klein sind. Es gilt

$$f_T = \frac{g_m}{2\pi C_{GG}}, \quad (3-13)$$

wobei die Kapazität C_{GG} folgendermaßen definiert ist:

$$C_{GG} = \left. \frac{dQ_{\text{Gate}}}{dV_{\text{GS}}} \right|_{V_{\text{DS}}=\text{const.}}. \quad (3-14)$$

Gleichung (3-13) entspricht Gleichung (2-22), da $C_{GG} = C_{GS} + C_{GD}$ [106]. Sie ist hier anwendbar, da die Serienwiderstände R_S und R_D des untersuchten DG MOSFET vernachlässigbar klein sind. Indem man eine Verzögerungszeit τ_{SD} einführt, lässt sich Gleichung (3-13) auch schreiben als

$$f_T = \frac{1}{2\pi\tau_{SD}}, \quad (3-15)$$

wobei unter Berücksichtigung der Definitionen (2-4) und (3-14) gilt

$$\tau_{SD} = \frac{dQ_{\text{Gate}}}{dI_D}. \quad (3-16)$$

Die Ladung auf der Gateelektrode kann durch ihre Gegenladung, die Inversionsladung im Kanal, ausgedrückt werden.

$$Q_{\text{Gate}} = q \int_0^L n_s(x) dx \quad (3-17)$$

Hierbei ist Q_{Gate} auf die Gateweite normiert, L ist die Kanallänge und x die Koordinate in Transportrichtung. Für I_D (normiert auf die Gateweite) gilt ohne Einschränkung

$$I_D = q \cdot n_s(x) \cdot \langle v_{||}(x) \rangle, \quad (3-18)$$

wobei $\langle v_{||}(x) \rangle$ die mittlere Elektronengeschwindigkeit in x -Richtung am Ort x ist. Mit Gleichung (3-17) und (3-18) wird aus Gleichung (3-16)

$$\tau_{\text{SD}} = q \int_0^L \frac{dn_s(x)}{dI_D} dx = \int_0^L \gamma(x) dx \quad (3-19)$$

wobei

$$\gamma(x) = q \frac{dn_s(x)}{dI_D}. \quad (3-20)$$

Die Funktion $\gamma(x)$ hat die Eigenschaften einer reziproken Geschwindigkeit. Liegen Simulationsergebnisse für zwei benachbarte Gatespannungen V_{GS1} und V_{GS2} vor, lässt sich $\gamma(x)$ näherungsweise mit Hilfe von

$$\gamma(x) \approx q \frac{\Delta n_s(x)}{\Delta I_D} = \frac{n_{\text{S2}} - n_{\text{S1}}}{n_{\text{S2}} \langle v_{||2} \rangle - n_{\text{S1}} \langle v_{||1} \rangle} \Big|_x \quad (3-21)$$

bestimmen. Da sich im MOSFET normalerweise das Geschwindigkeitsprofil im Kanal nur geringfügig mit V_{GS} ändert, d.h. $\langle v_{||1}(x) \rangle \approx \langle v_{||2}(x) \rangle \approx \langle v_{||}(x) \rangle$, lässt sich (3-21) noch wesentlich zu

$$\gamma(x) \approx \frac{1}{\langle v_{||}(x) \rangle} \quad (3-22)$$

vereinfachen. Damit hängt $\gamma(x)$ also nur noch von der mittleren Elektronengeschwindigkeit am Ort x ab und es genügt bereits die Simulation eines Arbeitspunktes. In diesem Fall wird das Integral (3-19) zu

$$\tau_{SD} = \frac{L}{\bar{v}}, \quad (3-23)$$

wobei \bar{v} die über den gesamten Kanal gemittelte Elektronengeschwindigkeit ist.

Die drei für den Vergleich verwendeten CP-Varianten, im Folgenden mit CP1, CP2 oder CP3 bezeichnet, unterscheiden sich in der Verwendung von (3-21) oder (3-22) zur Bestimmung von τ_{SD} und in den Integrationsgrenzen von (3-19).

- CP1: Gleichung (3-22), Integration über das undotierte Kanalgebiet direkt unter dem Gate.
- CP2: Gleichung (3-21), Integration über das undotierte Kanalgebiet direkt unter dem Gate.
- CP3: Gleichung (3-21), Integration über das gesamte Halbleitergebiet vom Source-Kontakt zum Drain-Kontakt.

Die mit S³A und den drei CP-Varianten ermittelten f_T des DG MOSFET für $V_{GS} = V_{DS} = 1V$ sind in Abbildung 3.11 zusammengefasst. Für CP2 und CP3 wurden zur Bestimmung von $\gamma(x)$ die Gleichstromarbeitspunkte mit $V_{GS1} = 0.8V$ und $V_{GS2} = 1V$ simuliert.

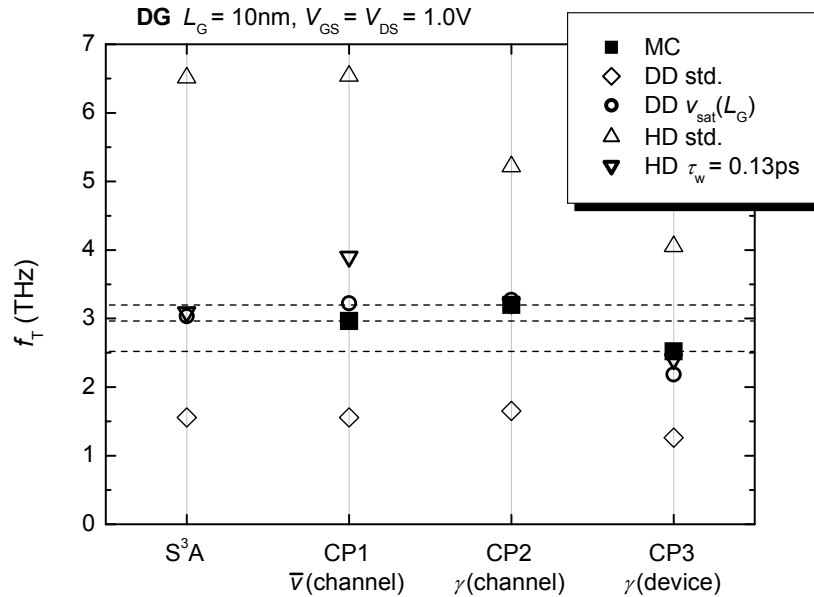


Abbildung 3.11 Mit dem DDM, dem HDM und mit MC simulierte f_T des untersuchten DG MOSFET. Für das DDM und das HDM wurden sowohl Standard- als auch modifizierte Parameter verwendet. Der Vergleich zeigt f_T -Werte, die mit S³A sowie mit drei Varianten der CP-Methode bestimmt wurden. Die gestrichelten horizontalen Linien markieren den Bereich der MC-Ergebnisse.

Auf den ersten Blick sieht Abbildung 3.11 etwas unübersichtlich aus, liefert aber bei genauerem Hinsehen die wesentlichen Erkenntnisse. Betrachten wir zunächst nur die S^3A -Ergebnisse. Mit dieser Methode liefern die modifizierten DD- und HD-Modelle nahezu identische f_T von etwa 3THz^4 . Im Vergleich dazu ist das f_T aus dem Standard-DDM nur etwa halb so groß und das aus dem Standard-HDM mehr als doppelt so groß. Vergleicht man die mit S^3A und die mit CP ermittelten f_T des DDM, stellt man fest, dass alle Methoden für die jeweilige Variante des DDM sehr ähnliche Ergebnisse liefern. Die CP-Werte des modifizierten DDM stimmen darüber hinaus sehr gut mit den entsprechenden MC-Ergebnissen überein. Das modifizierte DDM scheint demnach nicht nur zur Simulation des Gleichstromverhaltens geeignet zu sein, sondern beschreibt auch das Kleinsignalverhalten relativ gut.

Anders ist die Situation beim HDM. Hier ist die Streuung der mit den verschiedenen Methoden bestimmten f_T -Werte deutlich größer, besonders extrem beim HDM mit Standard-Parametern. Im Vergleich zu den MC-Ergebnissen sind die f_T -Werte des Standard-HDM durchweg viel zu groß. Für das angepasste HDM liefert CP1 ein deutlich größeres f_T als für MC. Dieses Ergebnis war zu erwarten, da auch das angepasste HDM zu hohe Elektronengeschwindigkeiten im Kanal berechnet (vgl. Abbildung 3.8). Es ist daher erstaunlich, dass mit CP2 und CP3 die Übereinstimmung zwischen angepasstem HDM und MC trotzdem ausgesprochen gut ist.

Abbildung 3.12 zeigt die mit dem DDM, dem HDM und mit MC berechneten $\gamma(x)$ entlang des Kanals. Darin ist zu sehen, dass die Funktion $\gamma(x)$ aus dem angepassten HDM ähnlich gut mit den MC-Ergebnissen übereinstimmt, wie die aus dem modifizierten DDM. Dagegen liefert das Standard DDM deutlich zu hohe Werte für $\gamma(x)$ und das Standard HDM im Schnitt deutlich zu kleine.

Insgesamt zeigt der Vergleich, dass das modifizierte DDM von den infrage kommenden Simulationsmodellen am besten zur Simulation der Hochfrequenzeigenschaften der in dieser Arbeit untersuchten MOSFET-Strukturen geeignet ist. Es liefert mit allen Berechnungsmethoden vergleichbare f_T , welche darüber hinaus gut mit den MC-Ergebnissen übereinstimmen. Für Kleinsignalanalysen wird im Weiteren die S^3A -Methode verwendet.

⁴ Derart hohe f_T erscheinen zwar aus heutiger Sicht unrealistisch für Silizium MOSFETs, werden aber auch von anderen Gruppen (z.B. [110], [111]) simuliert. Man muss beachten, dass es sich hierbei um das intrinsische f_T des MOSFET handelt, ohne Berücksichtigung parasitärer Widerstände und Kapazitäten.

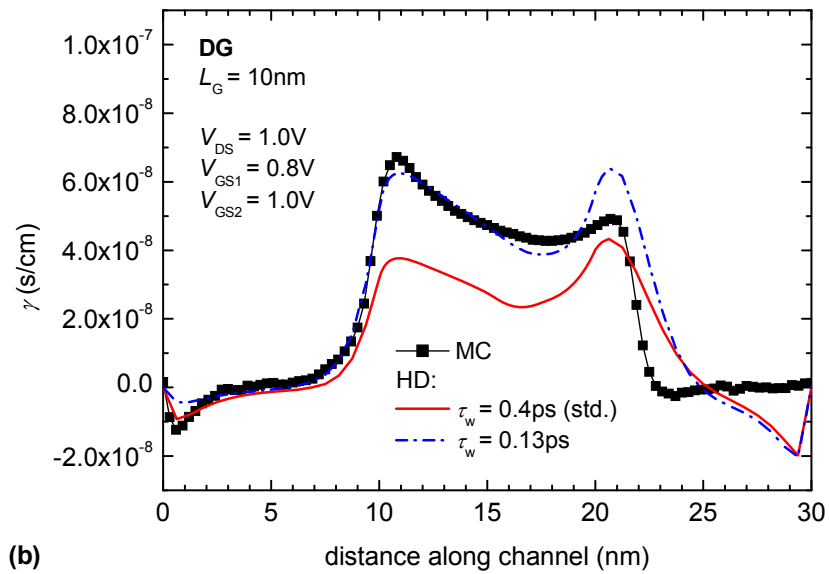
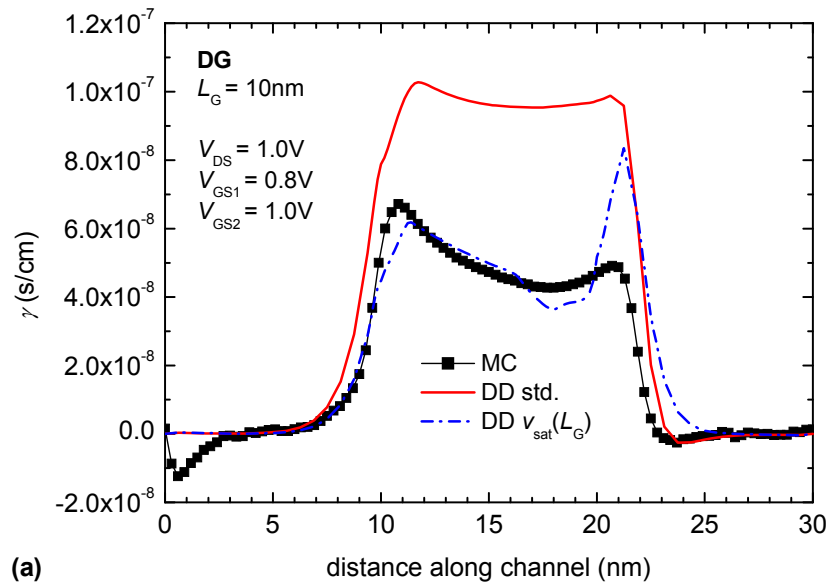


Abbildung 3.12 Nach Gleichung (3-21) bestimmte Funktion $\gamma(x)$ entlang des Kanals des DG MOSFET.
 (a) Vergleich DDM – MC. (b) Vergleich HDM – MC.

3.4. Quanteneffekte

Zu den Quanteneffekten in extrem skalierten MOSFETs gehören Tunneleffekte sowie die Formierung von Subbändern infolge von Quantum Confinement. Parasitäre Tunnelströme können, wie bereits in Abschnitt 2.4.3. besprochen, durch geeignete Designmaßnahmen in akzeptablen Grenzen gehalten werden. Band-zu-Band Tunneln zum Beispiel, das für die Skalierung des konventionellen Bulk-MOSFET eine fundamentale Grenze darstellt, spielt in MOSFET-Konzepten mit undotiertem Kanal kaum eine Rolle. Tunneln durch das Gateoxid lässt sich durch die Verwendung von high-K Material als Gateisolator wirksam unterdrücken und direktes Tunneln von Source nach Drain ist erst bei Gatelängen deutlich unterhalb von 10nm von Bedeutung. Unvermeidbar und nicht vernachlässigbar sind jedoch die Auswirkungen des Quantum Confinement. Dazu zählen im Wesentlichen eine Verschiebung der Transferkennlinie, d.h. eine Erhöhung der Schwellspannung im Vergleich zur klassischen Erwartung sowie eine Verringerung der Gatekapazität. Darüber hinaus hat die Formierung von Subbändern Auswirkungen auf die Elektronenbeweglichkeit im Kanal. Diese drei Effekte werden im Folgenden mit Hilfe von analytischen Modellen und numerischen Simulationen untersucht und diskutiert.

3.4.1. Verschiebung der Schwellspannung

Dieser Effekt lässt sich sehr effektiv mit Hilfe analytischer Modelle beschreiben, die auf bekannten Lösungen der Schrödingergleichung für bestimmte Potentialverläufe basieren. Derartige Modelle gibt es bereits für eindimensionale Effekte in Bulk-MOSFETs [29], in symmetrischen DG-MOSFETs [112] sowie in unsymmetrischen DG und SG MOSFETs [65]. In MG MOSFETs mit geringem Kanalquerschnitt, man spricht auch von Nanowire (NW) -MOSFETs, ist aufgrund des zweidimensionalen Quantum Confinement die Subbandformierung besonders stark ausgeprägt. Für MG MOSFETs mit rechteckigem Querschnitt wurde im Rahmen dieser Arbeit ein geeignetes Modell für die Schwellspannungs-verschiebung entwickelt [66]. Es folgt eine Herleitung dieses Modells und im Anschluss daran eine Diskussion wichtiger Ergebnisse.

Abbildung 3.13 zeigt beispielhaft MG MOSFET Strukturen, auf die sich die folgenden Überlegungen beziehen, sowie das verwendete Koordinatensystem. Näherungsweise lässt sich der Kanal dieser Transistoren als zweidimensionaler rechteckiger Potentialtopf mit unendlich hohen Wänden beschreiben. Bei undotiertem Kanal ist das

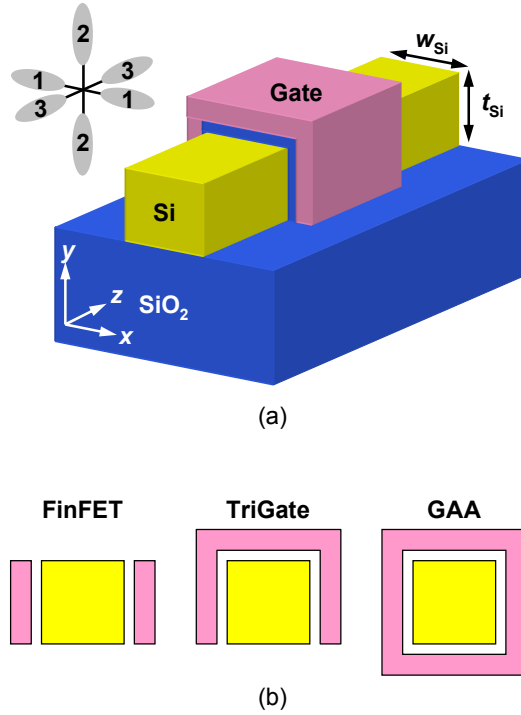


Abbildung 3.13 MG MOSFETs mit rechteckigem Querschnitt. (a) Prinzipielle Struktur eines Tri-Gate MOSFET mit der Definition der Kanalabmessungen w_{Si} und t_{Si} . Skizziert ist außerdem die Ausrichtung der Äquienergieflächen in den sechs Leitbandminima von Silizium für den Fall, dass alle Oberflächen des Kanals (100)-Orientierung haben. (b) Schematische Querschnitte dreier wichtiger MG MOSFET-Varianten.

Kanalpotential im Subthresholdbereich nahezu konstant [65], [113], so dass der Boden des Potentialtopfes als flach angesehen werden kann. Die Schrödingergleichung in Effektiver-Masse-Näherung für diese Anordnung lässt sich nach einigen Substitutionen [114]-[115] folgendermaßen aufschreiben

$$-\frac{\hbar}{2} \left(\frac{1}{m_{xx}} \frac{\partial^2 \psi}{\partial x^2} + \frac{1}{m_{yy}} \frac{\partial^2 \psi}{\partial y^2} + \frac{2}{m_{xy}} \frac{\partial^2 \psi}{\partial x \partial y} \right) - (V(x, y) + E) \psi = 0. \quad (3-24)$$

In Gleichung (3-24) sind $1/m_{xx}$, $1/m_{yy}$ und $1/m_{xy}$ Elemente des Tensors der reziproken effektiven Masse, $\psi(x, y)$ ist die Einhüllende der Wellenfunktion für den Energieeigenwert E und $V(x, y)$ das Potential⁵ des Potentialtopfes, wobei

$$V(x, y) = \begin{cases} 0 & \text{für } 0 < x < w_{Si} \text{ und } 0 < y < t_{Si} \\ \infty & \text{sonst} \end{cases}. \quad (3-25)$$

⁵ hier: potentielle Energie

Gleichung (3-24) muss für jedes der sechs Leitbandminima in Silizium gelöst werden. Abbildung 3.13 zeigt die Ausrichtung der Äquienergieflächen in den sechs Minima für den Fall, dass alle Oberflächen des Kanals (100)-Orientierung besitzen. Es ergeben sich 3 Paare äquivalenter Minima, die in Abbildung 3.13 entsprechend nummeriert sind. Für alle Minima gilt im betrachteten Fall $1/m_{xy} = 0$ und aus (3-24) erhält man die Energieeigenwerte

$$E_{k,mn} = \frac{h^2}{8} \left(\frac{m^2}{m_{xk}^* w_{Si}^2} + \frac{n^2}{m_{yk}^* t_{Si}^2} \right). \quad (3-26)$$

Die $E_{k,mn}$ sind nichts anderes als die energetischen Abstände zwischen Leitbandkante und den Bandkanten der einzelnen Subbänder. Der Index k kennzeichnet das Paar äquivalenter Leitbandminima entsprechend Abbildung 3.13, m und n sind die Quantenzahlen ($m, n = 1, 2, 3, \dots$) und m_{xk}^* und m_{yk}^* die effektiven Massen im k -ten Minimum in x - bzw. y -Richtung (m_{xx} bzw. m_{yy} in (3-24)).

Unter Subthreshold-Bedingungen ist die Verwendung der Boltzmann-Statistik gerechtfertigt. Damit ergibt sich der Beitrag eines Subbandes zur Linienkonzentration von Elektronen im Kanal zu

$$n_{k,mn}^{1D} = N_k^{1D} \exp\left(\frac{E_F - E_C^{QM}}{k_B T}\right) \exp\left(-\frac{E_{k,mn}}{k_B T}\right), \quad (3-27)$$

wobei E_F die Fermienergie ist, E_C^{QM} die quantenmechanisch berechnete Leitbandkante und

$$N_k^{1D} = \nu_k \left(\frac{2m_{dk}^* k_B T}{\pi \hbar^2} \right)^{1/2} \quad (3-28)$$

die eindimensionale effektive Zustandsdichte [116] im Minimum k mit dem Entartungsfaktor ν_k und der effektiven Zustandsdichtemasse m_{dk}^* . Summiert man die Beiträge aller Subbänder auf, so ergibt sich die Linienkonzentration der Elektronen zu

$$n_{1D}^{QM} = \exp\left(\frac{E_F - E_C^{QM}}{k_B T}\right) \sum_k \sum_m \sum_n N_k^{1D} \exp\left(-\frac{E_{k,mn}}{k_B T}\right). \quad (3-29)$$

Bei rein klassischer Betrachtung erhält man für die Linienkonzentration

$$n_{1D}^{CL} = w_{Si} t_{Si} N_C \exp\left(\frac{E_F - E_C^{CL}}{k_B T}\right), \quad (3-30)$$

wobei N_C die effektive Zustandsdichte im Leitungsband von Bulk-Silizium ist. Gibt man eine Linienkonzentration vor, so erhält man aus der Gleichsetzung $n_{1D}^{QM} = n_{1D}^{CL}$ die Verschiebung des quantenmechanischen Kanalpotentials gegenüber dem klassischen Fall

$$\Delta\phi^{QM} = \frac{1}{q} (E_C^{CL} - E_C^{QM}) = \frac{k_B T}{q} \ln \left(\frac{w_{Si} t_{Si} N_C}{\alpha} \right) \quad (3-31)$$

mit

$$\alpha = \sum_k \sum_m \sum_n N_k^{1D} \exp \left(-\frac{E_{k,mn}}{k_B T} \right). \quad (3-32)$$

Berücksichtigt man einen im Allgemeinen nicht idealen Subthreshold Slope S des betrachteten MOSFET, so ergibt sich die Verschiebung der Schwellspannung durch Quanteneffekte zu [65]

$$\Delta V_{th}^{QM} = \frac{qS}{k_B T \ln 10} \Delta\phi^{QM}. \quad (3-33)$$

Eine klassisch berechnete Schwellspannung V_{th}^{CL} kann nun mit ΔV_{th}^{QM} korrigiert werden und die tatsächliche Schwellspannung unter Berücksichtigung von Quanteneffekten ist

$$V_{th} = V_{th}^{CL} + \Delta V_{th}^{QM}. \quad (3-34)$$

Ähnliche Überlegungen zum eindimensionalen Quantum Confinement in SG SOI ($w_{Si} \rightarrow \infty$, vgl. Abbildung 3.13) oder auch in DG MOSFETs ($t_{Si} \rightarrow \infty$) führen zu

$$\Delta\phi_{DG}^{QM} = \frac{k_B T}{q} \ln \left(\frac{w_{Si} N_C}{\beta} \right), \quad \beta = \sum_k \sum_n N_k^{2D} \exp \left(-\frac{E_{k,n}}{k_B T} \right), \quad (3-35)$$

$$N_k^{2D} = \nu_k \frac{m_{dk}^* k_B T}{\pi \hbar^2}, \quad E_{k,n} = \frac{\hbar^2 n^2}{8 m_{xk}^* w_{Si}^2},$$

wobei N_k^{2D} die zweidimensionale effektive Zustandsdichte ist und w_{Si} hier sowohl für DG als auch für SG MOSFETs ganz allgemein die Schichtdicke des Kanals angibt. In Tabelle 3.1 sind die effektiven Massen und Entartungsfaktoren für beide Modelle zusammengefasst. Wenn nicht anders angegeben, werden diese Werte in den folgenden Berechnungen verwendet.

	MG			SG, DG	
Valley, k	1	2	3	1	2
Degeneracy, ν_k	2	2	2	2	4
$m_{xk}^*[m_0]$	0.916	0.190	0.190	0.916	0.190
$m_{yk}^*[m_0]$	0.190	0.916	0.190	-	-
$m_{dk}^*[m_0]$	0.190	0.190	0.916	0.190	0.417

Tabelle 3.1 Effektive Massen, Entartungsfaktoren und Index äquivalenter Leitbandminima in Silizium für die Modellierung zweidimensionaler Quanteneffekte in MG MOSFETs sowie eindimensionaler Quanteneffekte in SG oder DG MOSFETs für den Fall, dass alle Oberflächen (100)-Orientierung besitzen. Die effektiven Massen sind in Einheiten der Ruhemasse freier Elektronen m_0 angegeben.

Um die Genauigkeit des MG-Modells bzw. des SG/DG-Modells zu testen, wurden mit (3-31) bzw. (3-35) die $\Delta\phi^{\text{QM}}$ für eine Reihe unterschiedlicher Querschnittsgeometrien berechnet. In Abbildung 3.14 werden die Ergebnisse mit Lösungen des eindimensionalen Schrödinger-Poisson-Solvers SCHRED [117] für SG und DG MOS-Strukturen verglichen, sowie mit $\Delta\phi^{\text{QM}}$ für Tri-Gate MOSFETs aus zweidimensionalen Schrödinger-Poisson (2D-SP) Simulationen. Letztere wurden mit Gleichung (3-33) aus den in [118] präsentierten V_{th} - und S -Werten berechnet. Die Bestimmung von $\Delta\phi^{\text{QM}}$ mit SCHRED geschah folgendermaßen: zunächst wurde für jede MOS-Struktur die Elektronenschichtkonzentration als Funktion der Gatespannung $n_s(V_G)$ klassisch und quantenmechanisch simuliert, daraus wurden mit Hilfe eines n_{Sth} -Kriteriums⁶ [119] die klassischen und quantenmechanischen Schwellspannungen extrahiert und $\Delta V_{\text{th}}^{\text{QM}}$ bestimmt, S wurde extrahiert und schließlich mit Gleichung (3-33) $\Delta\phi^{\text{QM}}$ berechnet.

Abbildung 3.14 zeigt die berechneten $\Delta\phi^{\text{QM}}$ als Funktion von w_{Si} mit t_{Si} als Parameter. Für zunehmende t_{Si} nähern sich die Ergebnisse des MG-Modells asymptotisch denen des SG/DG-Modells an. Letztere befinden sich in exzellenter Übereinstimmung mit den SCHRED-Ergebnissen und SCHRED liefert nahezu identische $\Delta\phi^{\text{QM}}$ für SG und DG, d.h. für die beiden Extremfälle der MG-Strukturen in Abbildung 3.13(b). Im Inset von Abbildung 3.14 werden die Ergebnisse des MG-Modells mit denen der 2D-SP Simulationen von Tri-Gate MOSFETs aus [118] verglichen. Die Berechnungen in [118]

⁶ n_{Sth} war $1.7 \times 10^{10} \text{ cm}^{-2}$ für SG sowie $3.4 \times 10^{10} \text{ cm}^{-2}$ für DG-Strukturen.

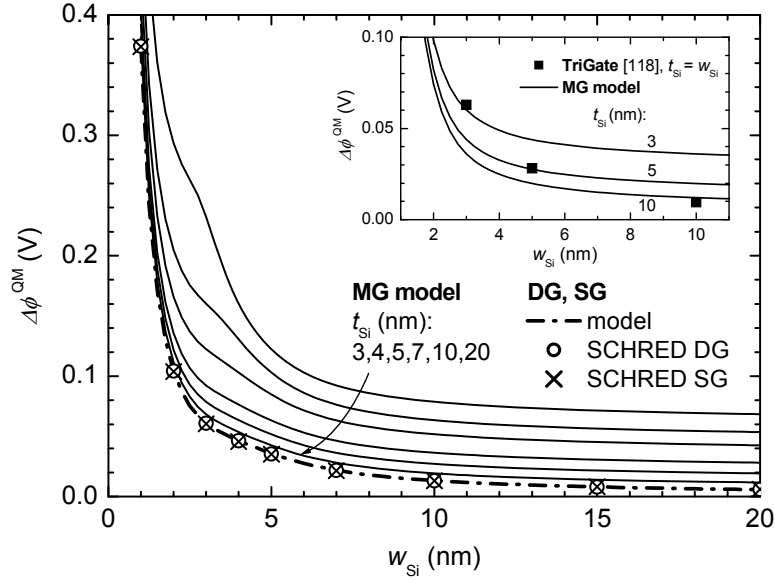


Abbildung 3.14 Mit dem MG- und dem SG/DG-Modell berechnete $\Delta\phi^{QM}$ als Funktion von w_{Si} im Vergleich zu SCHRED-Simulationen von SG und DG MOS-Strukturen. Im SCHRED-Input [117] wurden Metall-Gates sowie eine Gateoxiddicke von 1nm spezifiziert. Das BOX der SG SOI-Strukturen war 100nm dick. Für jeden Satz äquivalenter Minima wurde die Besetzung der jeweils 10 niedrigsten Subbänder berücksichtigt. Inset: Vergleich der $\Delta\phi^{QM}$ aus dem hier vorgestellten MG-Modell mit 2D-SP Ergebnissen für Tri-Gate MOSFETs [118]. Für diesen Vergleich wurde im MG-Modell die effektive Masse aus [118] verwendet.

basieren auf einem vereinfachten Modell für das Si-Leitungsband mit der effektiven Masse $m_{xk}^* = m_{yk}^* = m_{dk}^* = 1.08m_0$. Wird dieser Wert für die effektiven Massen innerhalb des MG-Modells verwendet, so werden damit die 2D-SP Ergebnisse für Tri-Gate MOSFETs sehr gut wiedergegeben.

In Abbildung 3.14 wird deutlich, dass ab $t_{Si} < 20\text{nm}$ $\Delta\phi^{QM}$ für MG-Strukturen deutlich größer ist als für die Grenzfälle SG und DG. Darüber hinaus lässt sich feststellen, dass der zweidimensionale Quanteneffekt stärker ist als die Summe zweier separat betrachteter eindimensionaler Effekte. Zum Beispiel ist das $\Delta\phi^{QM}$ für ein $4\text{nm} \times 4\text{nm}$ MG-Design um den Faktor 1.4 größer als das doppelte $\Delta\phi^{QM}$ einer DG-Struktur mit $w_{Si} = 4\text{nm}$.

Für Geometrien mit $t_{Si} \leq 5\text{nm}$ zeigen die Kurven $\Delta\phi^{QM}(w_{Si})$ eine Welle, deren Ursache in Abbildung 3.15 deutlich wird. Darin sind für ein MG-Design mit $t_{Si} = 4\text{nm}$ die jeweils niedrigsten Subbandenergien für die drei Paare äquivalenter Minima als

Funktion von w_{Si} gemeinsam mit $\Delta\phi^{\text{QM}}(w_{\text{Si}})$ dargestellt. Offensichtlich wird $\Delta\phi^{\text{QM}}$ vom jeweils niedrigsten $E_{k,mn}$ bestimmt. Bei $w_{\text{Si}} = t_{\text{Si}}$ wechselt die niedrigste Subbandenergie von $E_{1,11}$ auf $E_{2,11}$ wodurch die Welle in $\Delta\phi^{\text{QM}}(w_{\text{Si}})$ verursacht wird. Dieser Effekt ist im Inset von Abbildung 3.14 nicht zu sehen, da hier das Leitungsband von Silizium stark vereinfacht mit nur einer effektiven Masse, der Zustandsdichtemasse von Bulk-Si, beschrieben wird.

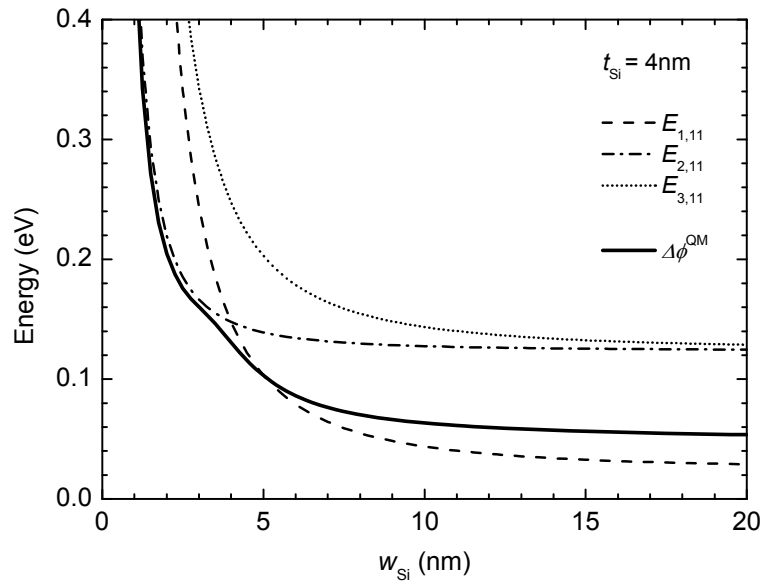


Abbildung 3.15 Mit Gleichung (3-26) berechnete niedrigste Subbandenergien für die drei Paare äquivalenter Minima sowie $\Delta\phi^{\text{QM}}$ als Funktion von w_{Si} für MG-Designs mit $t_{\text{Si}} = 4\text{nm}$.

Wie bereits in Abschnitt 2.4.3. diskutiert wurde, sind die hier untersuchten Quanteneffekte deshalb problematisch, weil durch sie die aufgrund von Fertigungstoleranzen stets vorhandenen V_{th} -Schwankungen extrem zunehmen, wenn die Abmessungen kleiner werden. Die maximale Abweichung von $\Delta\phi^{\text{QM}}$ von seinem Sollwert aufgrund von Schwankungen der Abmessungen w_{Si} und t_{Si} lässt sich mit Hilfe von

$$\delta\Delta\phi^{\text{QM}} = \left| \frac{\partial\Delta\phi^{\text{QM}}}{\partial w_{\text{Si}}} \Delta w_{\text{Si}} \right| + \left| \frac{\partial\Delta\phi^{\text{QM}}}{\partial t_{\text{Si}}} \Delta t_{\text{Si}} \right| \quad (3-36)$$

berechnen. Die partiellen Ableitungen von $\Delta\phi^{\text{QM}}$ für Gleichung (3-36) sind in Anhang B zusammengefasst. Abbildung 3.16 zeigt die Summe beider Ableitungen für

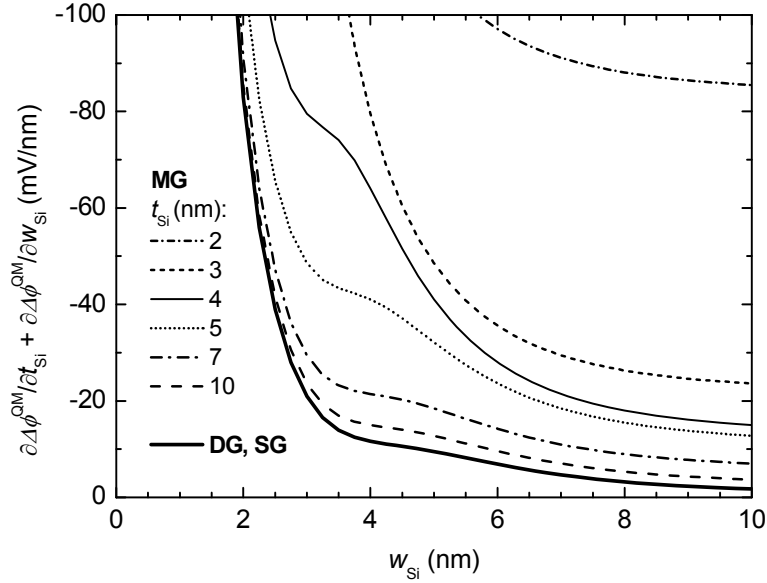


Abbildung 3.16 Vollständige Ableitung von $\Delta \phi^{QM}$ nach der Querschnittsgeometrie.

verschiedene Querschnittsgeometrien. Diese ist betragsmäßig gleich $\delta \Delta \phi^{QM}$ für den Fall $\Delta w_{Si} = \Delta t_{Si} = 1\text{nm}$. Nehmen wir als Beispiel ein maximal tolerierbares $\delta \Delta \phi^{QM}$ von 10mV. Dies entspricht einer Schwankung des Off-Stroms um den Faktor 1.5, die ausschließlich auf Quanteneffekte zurückzuführen ist⁷. In diesem Fall ist für einen SG oder DG MOSFET mit $w_{Si} = 4\text{nm}$ das maximal zulässige Δw kleiner als 1nm. Bei einem $4\text{nm} \times 4\text{nm}$ MG-Design dürfen die Geometrieschwankungen im Fall $\Delta w_{Si} = \Delta t_{Si}$ nicht größer sein als 1.5\AA . Das liegt im Bereich der Dicke einer Atomlage. Dieses Beispiel unterstreicht die enorme technologische Herausforderung für die Massenfertigung derartiger MOSFETs.

3.4.2. Einfluss von Quanteneffekten auf die Gatekapazität

Für die Schaltgeschwindigkeit eines MOSFET ist seine intrinsische Gatekapazität von besonderer Bedeutung. Das wird deutlich, wenn man die Gateverzögerungszeit (gate delay) τ mit Hilfe der CV/I-Metrik [Gleichung (2-6)] abschätzt. Dabei gibt $C = C_G + C_{\text{par}}$ die gesamte Gatekapazität des Transistors inklusive der intrinsischen Gatekapazität C_G sowie parasitärer Kapazitäten C_{par} an, $V = V_{\text{DD}}$ ist die Betriebsspannung und $I = I_{\text{on}}$

⁷ Wie Schwankungen des Off-Stromes mit Schwankungen von $\Delta \phi^{QM}$ zusammenhängen wird in Abschnitt 4.2.4. ausführlich behandelt.

ist der Drainstrom im eingeschalteten Zustand. Da I_{on} proportional zum Verhältnis C_G/L_G ist, hängt die Gateverzögerungszeit vom Faktor $(1 + C_{\text{par}}/C_G)$ ab. Dieser Faktor ist grundsätzlich größer als 1. Er kann jedoch minimiert werden, wenn C_G möglichst groß im Vergleich zu den parasitären Kapazitäten ist. Da in realen Bauelementen parasitäre Kapazitäten immer eine Rolle spielen, ist eine große flächenbezogene intrinsische Gatekapazität für eine hohe Schaltgeschwindigkeit von Vorteil.

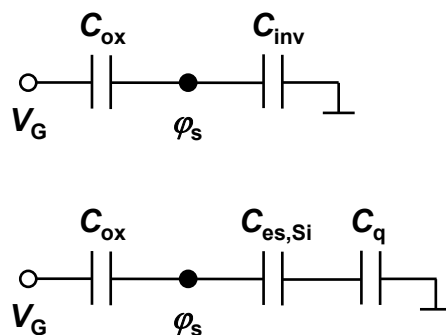


Abbildung 3.17 Ersatzschaltbild für die Gatekapazität von Silizium MOS-Strukturen. C_{ox} ist die Oxidkapazität und ϕ_s das Silizium-Oberflächenpotential. Die Inversionskapazität C_{inv} lässt sich in zwei Komponenten zerlegen, die elektrostatische Kapazität $C_{\text{es,Si}}$ sowie die Quantenkapazität C_q .

Generell lässt sich die Gatekapazität eines MOSFET mit Hilfe der Ersatzschaltung in Abbildung 3.17, bestehend aus Oxidkapazität C_{ox} und Inversionskapazität C_{inv} , beschreiben [2], [120]. Letztere lässt sich in zwei Komponenten zerlegen, die elektrostatische Kapazität $C_{\text{es,Si}}$ sowie die Quantenkapazität C_q [121] der Inversionsschicht. Während $C_{\text{es,Si}}$ die kapazitive Wirkung der räumlichen Verteilung der Kanalelektronen beschreibt, hängt C_q mit der endlichen elektronischen Zustandsdichte (density of states, DOS) im Silizium zusammen. Es gilt

$$C_G = C_{\text{ox}} \cdot \frac{1}{1 + C_{\text{ox}}/C_{\text{inv}}} = \left(\frac{1}{C_{\text{ox}}} + \frac{1}{C_{\text{es,Si}}} + \frac{1}{C_q} \right)^{-1}. \quad (3-37)$$

Im Idealfall, d.h. wenn C_{inv} viel größer ist als C_{ox} , erreicht C_G annähernd die Größe der Oxidkapazität. Das ist jedoch nur der Fall für relativ dicke Oxide in der Größenordnung von 10nm. Die Skalierung der Oxiddicke in den Bereich von 1nm hat zur Folge, dass C_G deutlich kleiner wird als C_{ox} [120]-[122]. Das ist zum Teil auf Quantum Confinement zurückzuführen, wodurch der Ladungsschwerpunkt der Kanalelektronen im Vergleich zu klassischen Betrachtungen weiter von der Si/SiO₂-Grenzfläche

weggedrängt wird [123]. Je größer der mittlere Abstand der Kanalelektronen von der Grenzfläche ist, desto kleiner ist $C_{\text{es,Si}}$. Wird zur Unterdrückung von Kurzkanaleffekten die Kanaldicke reduziert, z.B. durch Reduzierung der Siliziumdicke von SOI MOSFETs, verstärkt sich die Aufspaltung des Leitungsbandes in Subbänder, und die Zustandsdichte für die Kanalelektronen wird kleiner. Dadurch reduziert sich die Quantenkapazität [124], was zusätzlich zur Verringerung von C_G beiträgt.

Betrachtet man auf Siliziumnanodrähten (SiNWs) basierende MOSFET-Konzepte, wie z.B. Tri-Gate [16]-[18] oder Gate-All-Around (GAA) [22]-[23] Strukturen, so hat man es mit eindimensionalen (1D) Elektronensystemen zu tun, da im Nanodraht die Bewegung der Elektronen in zwei Richtungen eingeschränkt ist. Die Quantenkapazität in SiNW MOSFETs hängt von der 1D Zustandsdichte (1D DOS) ab [125], welche im Unterschied zur 2D DOS innerhalb eines Subbandes stark mit zunehmender Elektronenenergie abnimmt.

Ist bei sehr starker Subbandaufspaltung nur das niedrigste Subband mit Elektronen besetzt, erreicht C_q seinen Minimalwert. Bei $C_{\text{ox}} > C_q$ wird das sogenannte Quantum-Capacitance-Limit (QCL) erreicht [126]-[127]. Es wurde argumentiert, dass MOSFETs im QCL Vorteile hinsichtlich des Power-Delay-Produkts haben können [127]. Das ändert jedoch nichts an der Feststellung, dass für eine minimale Gateverzögerungszeit das Verhältnis C_G/C_{par} auf jeden Fall möglichst groß sein muss.

In einer aktuellen theoretischen Studie zu zylindrischen GAA MOS-Strukturen wurde eine kontinuierliche Abnahme von C_G relativ zu C_{ox} festgestellt, wenn der Durchmesser des Nanodrahts auf 5nm verkleinert wird [128]. Diese Studie basiert jedoch auf einer vereinfachten Beschreibung der Bandstruktur von Silizium mit nur einer isotropen effektiven Masse [129]. Andere Studien verwenden eine rigorose Beschreibung der Siliziumbandstruktur [63], [130]-[131], betrachten jedoch nicht explizit die Gatekapazität von SiNW MOSFETs.

Im Rahmen dieser Arbeit wurden Skalierungseffekte auf die Gatekapazität von Tri-Gate MOSFETs mit Hilfe analytischer Modelle und numerischer 2D Schrödinger-Poisson Simulationen untersucht [132]. Im folgenden Abschnitt werden die theoretischen Grundlagen erläutert und die untersuchten Strukturen sowie der verwendete Schrödinger-Poisson-Solver vorgestellt. Im Anschluss daran werden die wichtigsten Ergebnisse der Untersuchungen zur Quantenkapazität und zur Gatekapazität diskutiert. Durch einen Vergleich von klassischen und quantenmechanischen Berechnungen werden die Effekte des Quantum Confinement auf C_G gezeigt. Den Abschluss bildet eine kurze Zusammenfassung.

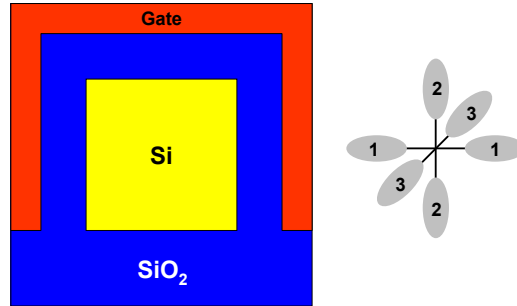


Abbildung 3.18 Querschnitt der untersuchten Tri-Gate MOS-Struktur. Das Gateoxid ist 1nm dick. Auf der rechten Seite sind die drei Paare äquivalenter Leitungsbandminima in Silizium für den Fall, dass alle Oberflächen (100)-Orientierung aufweisen, skizziert.

Theorie

Die hier untersuchte Tri-Gate MOS-Struktur ist in Abbildung 3.18 skizziert. Wir betrachten in dieser Studie einen quadratischen Siliziumquerschnitt, ein 1nm dickes Gateoxid, 30nm vergrabenes Oxid und ein Metallgate. Mit Hilfe von NEXTNANO [133]-[134] wurden für diese Struktur die 2D Schrödinger- und Poissongleichungen selbstkonsistent gelöst. Die Schrödingergleichung wurde für jedes der 3 Paare äquivalenter Leitungsbandminima (vgl. Abb. 3.18) in der Effektiven-Masse-Näherung gelöst. Für die Si/SiO₂-Grenzfläche wurde eine endliche Barriere von 3.1eV angenommen, wodurch ein Eindringen der Wellenfunktionen in das Oxid zugelassen wird. Abhängig von den Abmessungen des Siliziumquerschnitts wurden bis zu 25 Subbänder pro Satz äquivalenter Minima berücksichtigt. Es wurden Strukturen mit verschiedenen Siliziumabmessungen simuliert, wobei jeweils die Gatespannung V_G variiert wurde. Für jeden Arbeitspunkt wurde die Linienkonzentration der Elektronen n_{1D} durch Integration der Elektronendichte über den gesamten Querschnitt der Tri-Gate MOS-Struktur bestimmt. Die Gatekapazität pro Längeneinheit erhält man dann aus der Ableitung

$$C_G = q \frac{\partial n_{1D}}{\partial V_G}. \quad (3-38)$$

Die Oxidkapazität einer Tri-Gate-Struktur mit quadratischem Querschnitt lässt sich mit Hilfe von

$$C_{ox} = \frac{N \cdot F \cdot \epsilon_{ox}}{\ln \left(1 + F \frac{t_{ox}}{w_{Si}} \right)} \quad (3-39)$$

modellieren, wobei N und F Fitparameter sind, $N = 3.05$ und $F = 1.5$ [131]. Sind C_G und C_{ox} bekannt, lässt sich die Inversionskapazität leicht über $C_{inv} = (1/C_G - 1/C_{ox})^{-1}$ bestimmen. Um zu einer Gleichung für die Quantenkapazität zu gelangen, gehen wir am besten von der Definition von C_{inv} aus:

$$C_{inv} = q \frac{\partial n_{1D}}{\partial \varphi_s}. \quad (3-40)$$

Das Oberflächenpotential φ_s ist definiert als $\varphi_s = -(E_{C,s} - E_F)/q$, wobei $E_{C,s}$ die Leitbandkante von Silizium an der Si/SiO₂-Grenzfläche ist. In einem 1D Elektronensystem ist die Linienkonzentration gegeben durch

$$n_{1D} = \sum_k \sum_m \sum_n \int_0^\infty g_{1D}^k(\varepsilon) f(\varepsilon, E_{k,mn}) d\varepsilon, \quad (3-41)$$

mit der von der Energie ε abhängigen 1D Zustandsdichte (1D DOS)

$$g_{1D}^k(\varepsilon) = \frac{v_k}{\pi \hbar} \sqrt{\frac{2m_{dk}^*}{\varepsilon}}, \quad (3-42)$$

und der Fermi-Dirac-Verteilung

$$f(\varepsilon, E_{k,mn}) = \frac{1}{1 + \exp\left(\frac{\varepsilon + E_{k,mn} - q\varphi_s}{k_B T}\right)}. \quad (3-43)$$

Entsprechend (3-40) erhält man durch Differentiation von Gleichung (3-41) den Ausdruck

$$C_{inv} = q \sum_k \sum_m \sum_n \left(\int_0^\infty g_{1D}^k(\varepsilon) \frac{\partial}{\partial \varphi_s} f(\varepsilon, E_{k,mn}) d\varepsilon \right) \left(1 - \frac{\partial E_{k,mn}}{\partial \varphi_s} \right). \quad (3-44)$$

Die Inversionskapazität stellt sich demnach als eine Summe von Beiträgen der einzelnen Subbänder dar. Der erste Faktor in der Summe ist die Quantenkapazität eines Subbandes [125]. Der zweite Faktor $(1 - \partial E_{k,mn} / \partial \varphi_s)$ berücksichtigt den Effekt der räumlichen Verteilung der Elektronen im Kanal [124], [135], d.h. den Beitrag des Subbandes zu $C_{es,Si}$. Die gesamte Quantenkapazität C_q ergibt sich aus der Summe der Beiträge aller Subbänder. Nach der Ableitung von $f(\varepsilon, E_{k,mn})$ erhalten wir für C_q den Ausdruck

$$C_q = \frac{q^2}{\sqrt{2} h k_B T} \sum_k \sum_m \sum_n \nu_k \sqrt{m_{dk}^*} \int_0^\infty \frac{1}{\sqrt{\varepsilon}} \cosh^{-2} \left(\frac{\varepsilon + E_{k,mn} - q\varphi_s}{k_B T} \right) d\varepsilon. \quad (3-45)$$

Um mit Gleichung (3-45) die Quantenkapazität einer Tri-Gate-Struktur berechnen zu können, müssen die energetischen Abstände der Subbandkanten vom Fermi-niveau, $E_{k,mn} - q\varphi_s$, bekannt sein. Diese lassen sich mit Hilfe von analytischen Näherungslösungen oder durch rigorose selbstkonsistente 2D Schrödinger-Poisson-Simulationen bestimmen.

Um den Einfluss der Subbandformierung auf C_q quantifizieren zu können, wäre ein Vergleich der Quantenkapazität eines 1D Elektronensystems, gegeben durch Gleichung (3-45), mit der von Bulk-Silizium, d.h. eines 3D Elektronensystems, mit identischer Geometrie hilfreich. Vernachlässigt man die Bandverbiegung im Silizium, lässt sich die Quantenkapazität des 3D Elektronensystems mit Hilfe von

$$C_q^{\text{bulk}} = \frac{\nu q^2 t_{\text{Si}} w_{\text{Si}} m_d^* \sqrt{2m_d^*}}{4k_B T \pi^2 \hbar^3} \int_0^\infty \sqrt{\varepsilon} \cdot \cosh^{-2} \left(\frac{\varepsilon + E_C - E_F}{k_B T} \right) d\varepsilon \quad (3-46)$$

berechnen, wobei $\nu = 6$ der Entartungsfaktor in Bulk-Silizium ist und m_d^* die effektive Zustandsdichtemasse in Bulk-Silizium. Eine Herleitung von Gleichung (3-46) ist in Anhang C zu finden. Für einen fairen Vergleich müssen wir auch für das 1D Elektronensystem flache Bänder annehmen. Das führt zu etwas anderen Subbandenergien im Vergleich zur exakten Lösung, die wesentlichen Aussagen zur Quantenkapazität bleiben jedoch gleich. Nehmen wir weiterhin eine unendlich hohe Barriere an der Si/SiO₂-Grenzfläche an, lassen sich die Subbandenergien leicht mit Gleichung (3-26) berechnen. Die mit (3-45) und (3-26) berechnete Quantenkapazität eines 1D Elektronensystems kann nun direkt mit C_q^{bulk} aus Gleichung (3-46) verglichen werden.

Quantenkapazität

Beginnen wir die Diskussion mit analytischen Betrachtungen. Abbildung 3.19 zeigt die Quantenkapazität einer 5nm × 5nm SiNW-Struktur, berechnet mit Gleichungen (3-45) und (3-26), in Abhängigkeit von der Lage des Fermi-niveaus. Die energetischen Positionen der Subbänder sind durch vertikale Linien markiert, die horizontale Linie zeigt die Oxidkapazität zum Vergleich. Wie zu sehen ist, beginnt C_q stark anzusteigen, wenn das Fermi-niveau in die Nähe des ersten Subbandes gelangt. Je mehr Subbänder

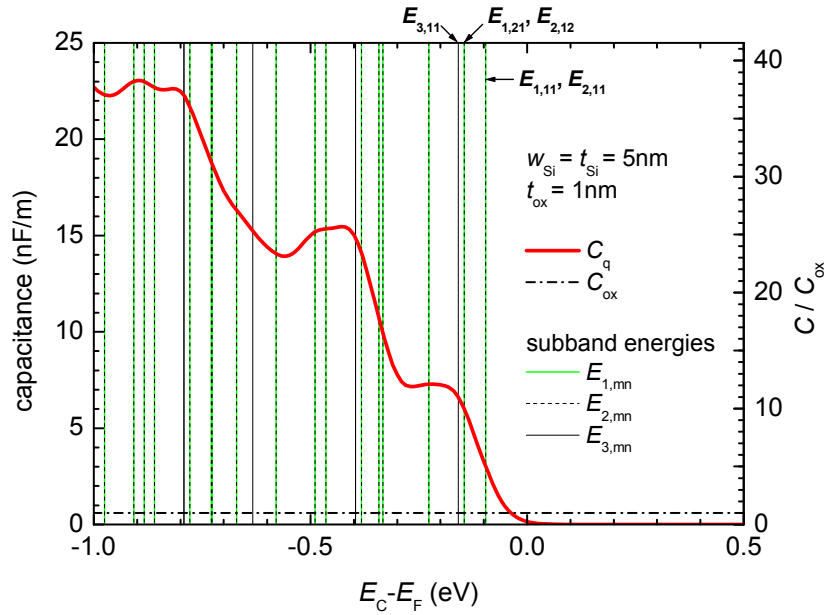


Abbildung 3.19 Quantenkapazität einer $5\text{nm} \times 5\text{nm}$ SiNW-Struktur in Abhängigkeit von der Lage des Ferminiveaus. Vertikale Linien markieren die energetische Lage der Subbänder nach Gleichung (3-26). Die horizontale Linie kennzeichnet die Oxidkapazität.

mit Elektronen besetzt werden, desto größer wird C_q . Hat E_F die ersten 4 Subbandkanten überquert (zweite vertikale Linie von rechts in Abb. 3.19), ist C_q bereits zehnmal größer als C_{ox} . Noch größere C_q wirken sich kaum noch auf die Gatekapazität aus.

Der Anstieg von C_q ist jedoch nicht kontinuierlich. Wird der Abstand der Subbänder zu groß, kann C_q wieder kleiner werden, selbst wenn bereits viele Subbänder besetzt sind. Der Grund dafür ist die Energieabhängigkeit der 1D DOS. Entsprechend Gleichung (3-42) nimmt g_{1D}^k mit zunehmender Energie ab, wodurch C_q kleiner wird. Das ist nicht der Fall in 3D Elektronensystemen [Gleichung (3-46)].

In Abbildung 3.20 werden die Quantenkapazitäten von 1D und 3D Elektronensystemen, C_q und C_q^{bulk} , für verschiedene Siliziumabmessungen miteinander verglichen. Beide Größen zeigen grundsätzlich den gleichen Trend: die Quantenkapazität sinkt mit abnehmenden Siliziumquerschnitt. In jedem Fall ist C_q^{bulk} jedoch größer als C_q , da aufgrund der Subbandformierung im 1D Fall die gesamte Elektronenzustandsdichte kleiner ist als in Bulk-Silizium. Der relative Unterschied zwischen C_q und C_q^{bulk} wird mit abnehmendem Querschnitt immer größer, da sich die Subbänder immer weiter

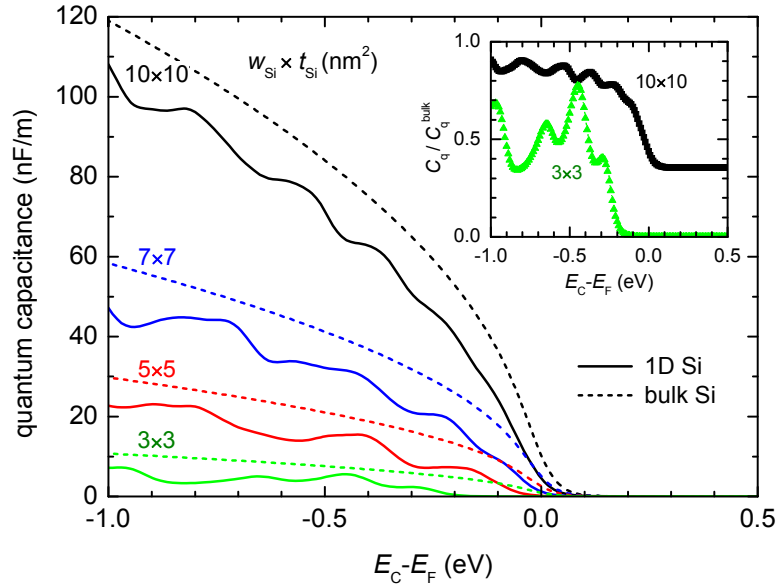


Abbildung 3.20 Quantenkapazität von Silizium-Nanodrähten mit verschiedenen Querschnitten. Quantenmechanische (durchgezogene Linien) und klassische (gestrichelte Linien) Ergebnisse im Vergleich.

voneinander entfernen (vgl. Inset von Abb. 3.20). Die Frage ist nun, wie stark die Gatekapazität durch C_q reduziert wird, wenn man die Tri-Gate-Struktur skaliert.

Abbildung 3.21 zeigt die Funktion $(1 + C_{ox}/C_q)^{-1}$, d.h. die Reihenschaltung von C_q und C_{ox} normiert auf C_{ox} , als Funktion der Elektronenlinienkonzentration für verschiedene Tri-Gate-Strukturen. In dieser Darstellung liefern beide Varianten, sowohl 1D als auch 3D Elektronensysteme, nahezu identische Kurven für $w_{Si} \times t_{Si} > 3\text{nm} \times 3\text{nm}$. Für kleinere Abmessungen beginnen die Kurven voneinander abzuweichen. Überraschenderweise kann jedoch die Gesamtkapazität für das 1D Elektronensystem größer sein als die für Bulk-Silizium. Das lässt sich mit Hilfe von Abbildung 3.22 verstehen, wo C_q und C_q^{bulk} als Funktion von n_{1D} dargestellt sind. Für kleine Abmessungen kann bei gleicher Linienkonzentration C_q größer sein als C_q^{bulk} . Das lässt sich mit der Energieabhängigkeit von g_{1D}^k [Gleichung (3-42)] erklären. Für $\varepsilon \rightarrow 0$, d.h. in Richtung Subbandkante, wird g_{1D}^k unendlich groß. Deshalb kann die Änderung der Elektronendichte im Vergleich zu Bulk-Silizium stärker sein, wenn E_F gerade die Subbandkante überquert.

In Abbildung 3.21 wurde für jede Struktur der besonders interessante Bereich markiert, in dem sich der Transistor nach den Vorgaben der ITRS im On-Zustand befinden soll.

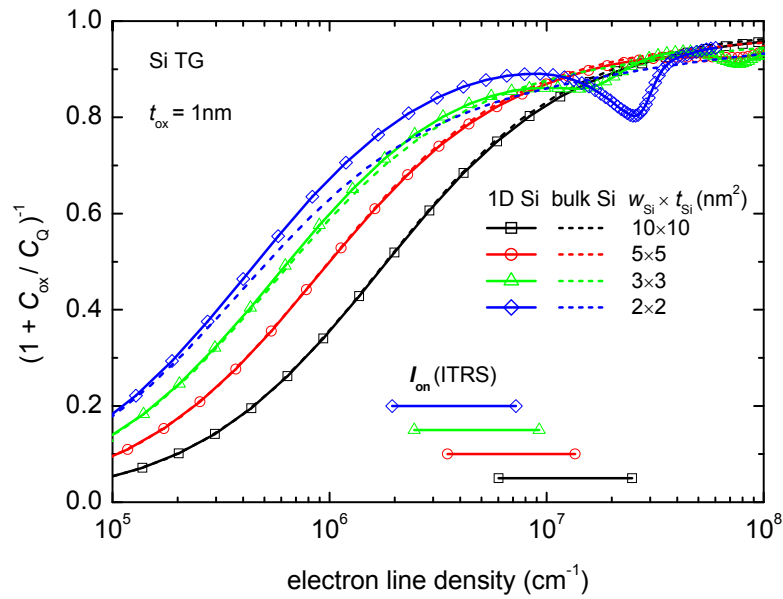


Abbildung 3.21 Reihenschaltung von C_{ox} und C_{q} als Funktion der Elektronenlinienkonzentration für Tri-Gate MOS-Strukturen mit verschiedenen Querschnitten. Quantenmechanische (durchgezogene Linien) und klassische (gestrichelte Linien) Ergebnisse im Vergleich. Horizontale Linien markieren für jede Struktur den Bereich von $n_{1\text{D}}$ im eingeschalteten Zustand.

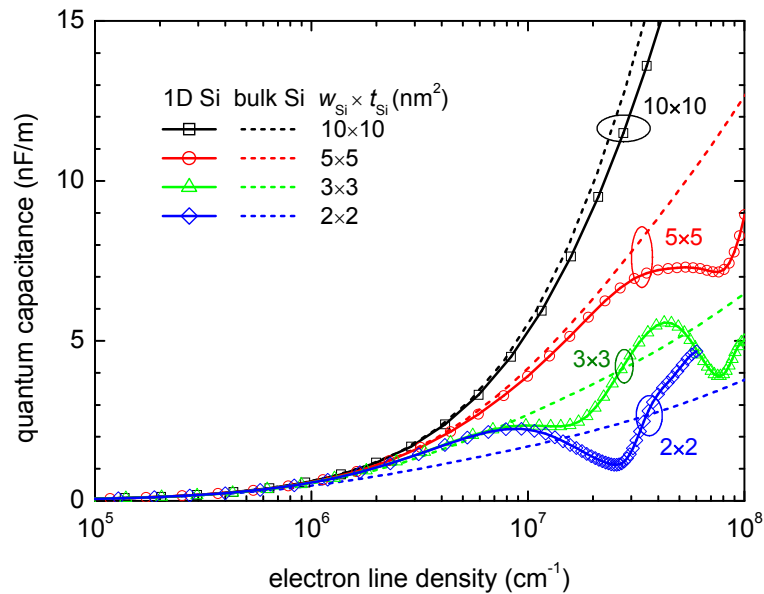


Abbildung 3.22 Quantenkapazität von Silizium-Nanodrähten mit verschiedenen Querschnitten als Funktion der Elektronenlinienkonzentration. Quantenmechanische (durchgezogene Linien) und klassische (gestrichelte Linien) Ergebnisse im Vergleich.

Dieser Bereich wurde mit Hilfe der in Abbildung 3.23 dargestellten $n_{1D}(V_G)$ -Kurven bestimmt, welche mit NEXTNANO simuliert wurden. Die vertikalen Linien grenzen den Gatespannungsbereich ein, für den in der aktuellen ITRS On-Ströme definiert sind. Je nach Logikfamilie und Jahr findet man I_{on} -Definitionen bei effektiven Gatespannungen, $V_G - V_{th}$, zwischen 0.28V und 0.72V [4]. Vergleicht man nun in Abbildung 3.21 die Kapazitäten verschiedener Strukturen aus den markierten Bereichen miteinander, so stellt man fest, dass sie in etwa gleich groß sind. Dieses Ergebnis ist auf den ersten Blick überraschend. Es wird jedoch verständlich, wenn man bedenkt, dass sowohl C_q als auch C_{ox} mit den Abmessungen des Nanodrahtes skalieren (siehe auch Abb. 3.26).

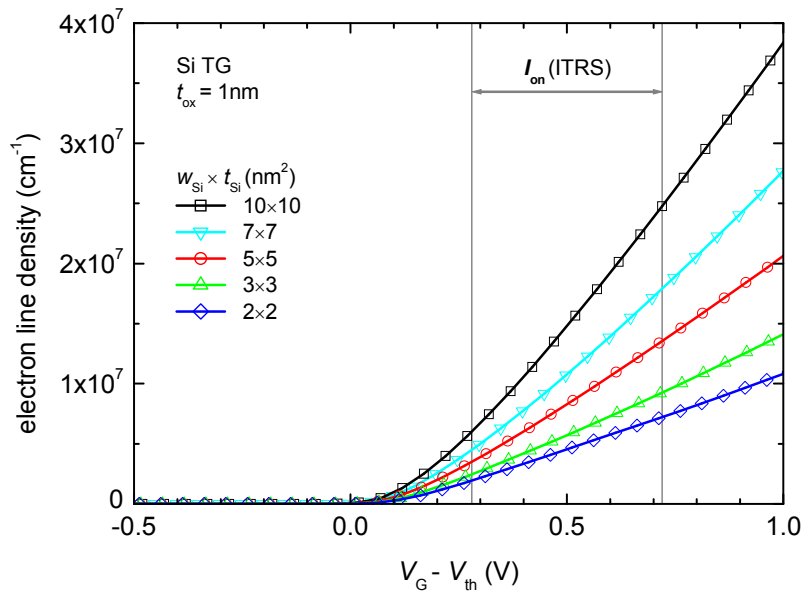


Abbildung 3.23 Elektronenlinienkonzentration in Tri-Gate MOS-Strukturen mit verschiedenen Querschnitten als Funktion der effektiven Gatespannung aus quantenmechanischen NEXTNANO-Simulationen. Vertikale Linien begrenzen den Gatespannungsbereich, für den in der aktuellen ITRS On-Ströme definiert sind.

Gatekapazität

In Abbildung 3.24 werden die mit klassischen und quantenmechanischen NEXTNANO-Simulationen bestimmten C_G für eine $5\text{nm} \times 5\text{nm}$ Tri-Gate-Struktur verglichen. Zum Vergleich sind noch einmal die entsprechenden Kurven aus Abbildung 3.21 für die Reihenschaltung von C_q (C_q^{bulk}) und C_{ox} gezeigt.

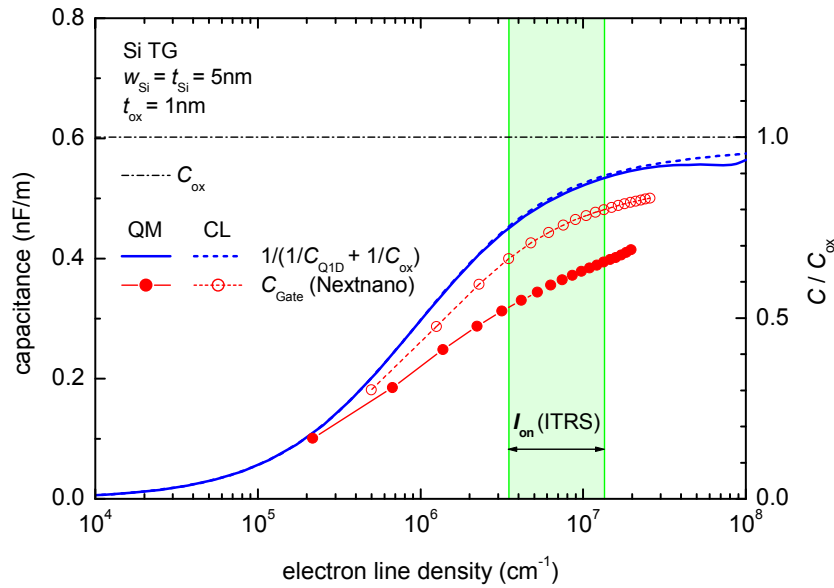


Abbildung 3.24 Gatekapazität und Reihenschaltung von C_{ox} und C_q als Funktion der Elektronenlinienkonzentration für eine $5\text{nm} \times 5\text{nm}$ Tri-Gate MOS-Struktur. Quantenmechanische (durchgezogene Linien bzw. volle Symbole) und klassische (gestrichelte Linien bzw. offene Symbole) Ergebnisse im Vergleich.

Vergleicht man C_G mit C_{ox} , lässt sich der Einfluss der endlichen Inversionskapazität auf C_G quantifizieren. Im On-Zustand erreicht C_G etwa 55-65% des Wertes von C_{ox} , während klassische Simulationen 65-80% von C_{ox} vorhersagen. Diese Abweichung ist hauptsächlich auf die Unterschiede in der Elektronendichteverteilung zwischen klassischen Lösungen und denen von Schrödinger-Poisson-Simulationen zurückzuführen. Unterschiede zwischen C_q und C_q^{bulk} haben, wie bereits diskutiert, nur einen geringen Einfluss auf die Gatekapazität. Der separate Effekt der Quantenkapazität auf C_G ist eine Reduzierung um 10-25% vom Idealwert C_{ox} .

Abbildung 3.25 zeigt die mit NEXTNANO simulierten quantenmechanischen und klassischen Elektronendichteverteilungen in einer $5\text{nm} \times 5\text{nm}$ Tri-Gate MOS-Struktur bei $V_G - V_{th} = 0.5\text{V}$. Im Vergleich zur quantenmechanischen Verteilung liegen die Elektronen bei klassischer Betrachtung viel dichter an der Si/SiO₂ Grenzfläche. Aus diesem Grund ist die elektrostatische Kapazität des Bulk-Siliziumkanals viel größer als das $C_{es,Si}$ des 1D Elektronensystems, was die geringere Gatekapazität für den zweiten Fall erklärt.

In Abbildung 3.26 ist zu sehen, wie die einzelnen Komponenten der Gatekapazität mit den Siliziumabmessungen skalieren. Sowohl in quantenmechanischen [Abb. 3.26(a)] als

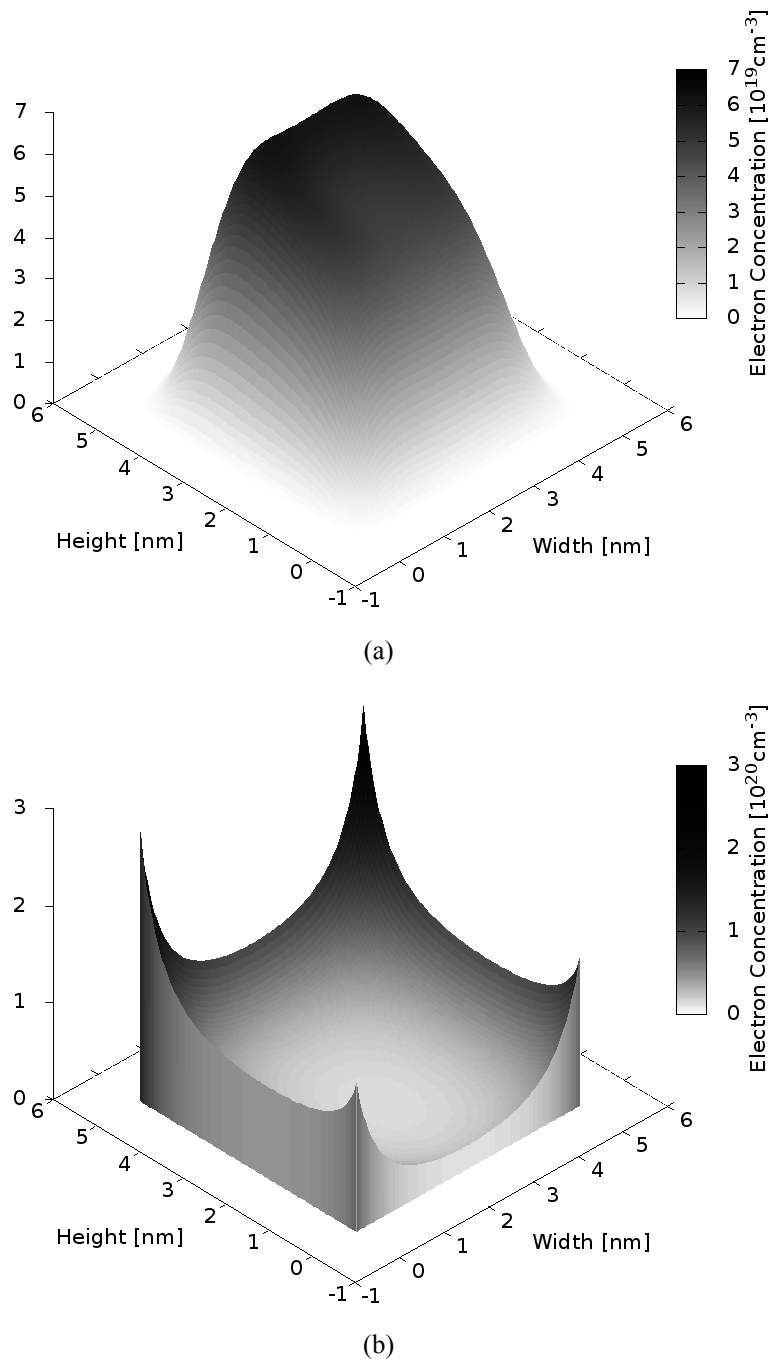


Abbildung 3.25 Mit NEXTNANO simulierte (a) quantenmechanische und (b) klassische Elektronendichteverteilung in einer $5\text{nm} \times 5\text{nm}$ Tri-Gate MOS-Struktur bei $V_G - V_{\text{th}} = 0.5\text{V}$.

auch in klassischen [Abb. 3.26(b)] Betrachtungen zeigen alle Komponenten eine nahezu lineare Abhängigkeit von w_{Si} bzw. t_{Si} . Ein ausgeprägt nichtlineares Verhalten lässt sich nur bei C_q für extrem skalierte SiNW-Strukturen mit $w_{\text{Si}} = t_{\text{Si}} < 3\text{nm}$ feststellen.

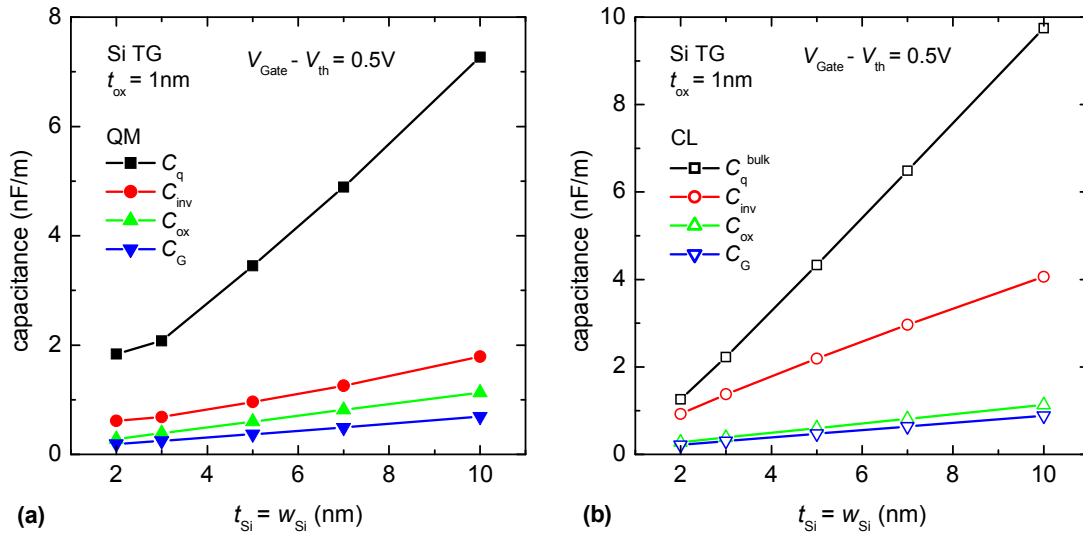


Abbildung 3.26 Die Komponenten der Gatekapazität von Trigate MOS-Strukturen als Funktion des Silizium-querschnitts. (a) Quantenmechanische und (b) klassische Ergebnisse.

Abbildung 3.27 zeigt die Gatekapazität normiert auf C_{ox} , d.h. auf die maximal mögliche Kapazität der entsprechenden Tri-Gate-Struktur, als Funktion der Siliziumabmessungen. Um den separaten Effekt der Quantenkapazität auf C_G zu veranschaulichen, ist die Funktion $(1 + C_{ox}/C_q)^{-1}$ ebenfalls gezeigt. Quantenmechanische und klassische Ergebnisse werden verglichen. In Abbildung 3.27 lässt sich eine kontinuierliche Vergrößerung des Verhältnisses C_G/C_{ox} feststellen, wenn der SiNW-Querschnitt kleiner als $7\text{ nm} \times 7\text{ nm}$ wird. Demgegenüber zeigen die klassischen Ergebnisse den entgegengesetzten Trend, auch wenn die klassischen C_G/C_{ox} -Werte nur um weniger als 3% variieren. Andererseits sagen die quantenmechanischen Berechnungen eine Vergrößerung von C_G/C_{ox} bis zu 14% voraus, wenn der Siliziumquerschnitt von $7\text{ nm} \times 7\text{ nm}$ auf $2\text{ nm} \times 2\text{ nm}$ reduziert wird. Das kann teilweise auf die Vergrößerung von C_q/C_{ox} zurückgeführt werden, jedoch entsprechend dem Verlauf von $(1 + C_{ox}/C_q)^{-1}$ nur für das $2\text{ nm} \times 2\text{ nm}$ Design. Von größerer Bedeutung ist die relative Zunahme von $C_{es,Si}$, da die mittlere Entfernung der Kanalelektronen von der Si/SiO₂-Grenzfläche abnimmt, wenn w_{Si} und t_{Si} skaliert werden.

Abbildung 3.28 veranschaulicht diesen Effekt. Hier sind in einem horizontalen Schnitt durch die Tri-Gate-Struktur sowohl die quantenmechanische als auch die klassische Elektronen-dichteverteilung für verschiedene Siliziumabmessungen dargestellt. Im klassischen Bild befindet sich das Dichtemaximum immer direkt an der Si/SiO₂

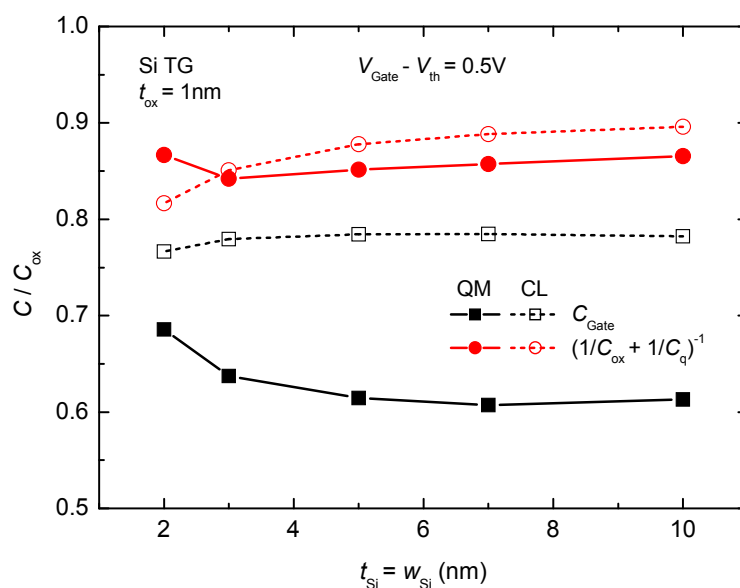


Abbildung 3.27 Gatekapazität und Reihenschaltung von C_{ox} und C_q normiert auf C_{ox} als Funktion der Siliziumabmessungen. Quantenmechanische (durchgezogene Linien) und klassische (gestrichelte Linien) Ergebnisse im Vergleich.

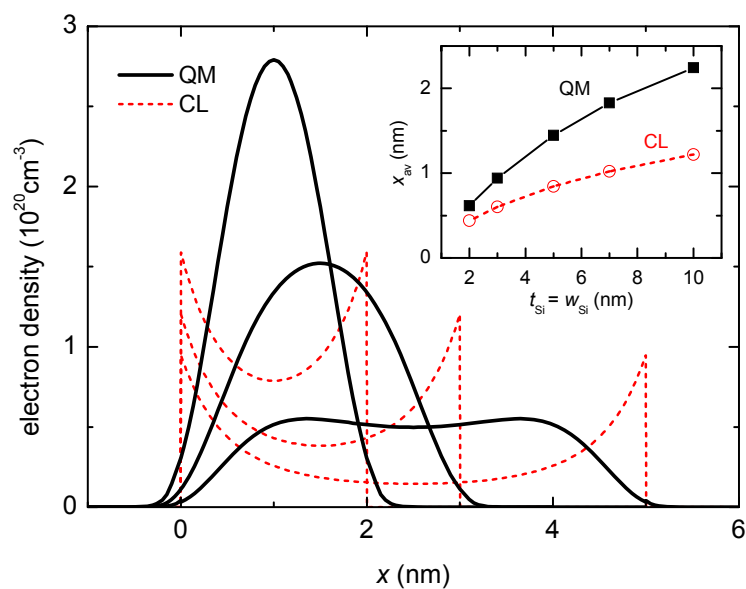


Abbildung 3.28 Quantenmechanische und klassische Elektronendichteverteilung in einem horizontalen Schnitt bei $t_{\text{Si}}/2$ durch die Tri-Gate-Struktur für verschiedene Siliziumabmessungen. Inset: mittlerer Abstand der Elektronen in der Schnittebene von der nächstgelegenen Grenzfläche als Funktion der Siliziumabmessungen.

Grenzfläche. Quantenmechanisch sind die Elektronen mehr im Innern des Siliziums verteilt. In der $5\text{nm} \times 5\text{nm}$ Struktur lassen sich gerade noch zwei separate, zur jeweiligen Grenzfläche zuordenbare Dichtemaxima erkennen. In noch kleineren Querschnitten verschmelzen beide Maxima zu einem einzigen und man spricht von Volumeninversion. In dieser Situation hängt die mittlere Entfernung eines Elektrons von der nächsten Grenzfläche, x_{av} , stark von Größe des Siliziumquerschnitts ab. Klassisch dagegen befinden sich die Elektronen in allen Siliziumquerschnitten sehr dicht an der Grenzfläche. Das bedeutet, wird der Querschnitt skaliert, dann verringert sich das klassisch berechnete x_{av} nicht so stark wie das quantenmechanisch berechnete (siehe Inset von Abb. 3.28).

Fazit

Die Berechnungen haben gezeigt, dass sich die Gatekapazität von Tri-Gate MOSFETs im Vergleich zur Oxidkapazität nicht verringert, wenn die Kanalabmessungen skaliert werden, obwohl die Quantenkapazität dabei stark abnimmt. Der Grund dafür ist, dass bei der Tri-Gate-Struktur sowohl C_{ox} als auch C_{q} mit dem Kanalquerschnitt skalieren. Das ist bei planaren MOSFET-Konzepten nicht der Fall, da die Oxidkapazität dort konstant bleibt. Tatsächlich sagen die Berechnungen für eine Skalierung des Kanalquerschnittes unterhalb von $7\text{nm} \times 7\text{nm}$ eine Zunahme des Verhältnisses $C_{\text{G}}/C_{\text{ox}}$ voraus. Diese Zunahme muss auf Quanteneffekte zurückgeführt werden, da klassische Simulationen keinen Anstieg von $C_{\text{G}}/C_{\text{ox}}$ liefern. Quanteneffekte beeinflussen die Gatekapazität in erster Linie über die räumliche Verteilung der Elektronen im Kanal. Effekte der Subbandformierung in der Quantenkapazität wirken sich kaum auf die Gatekapazität von Tri-Gate MOSFETs aus, außer bei extrem kleinen Siliziumquerschnitten im Bereich von $2\text{nm} \times 2\text{nm}$.

3.4.3. Elektronenbeweglichkeit in Silizium-Nanodrähten

Der Drainstrom in Nanometer-MOSFETs wird trotz ballistischer Transporteffekte immer noch wesentlich durch die effektive Elektronenbeweglichkeit μ_{eff} im Kanal bestimmt [98], [100]. Entsprechend Gleichung (3-7) hängt die mittlere Elektronengeschwindigkeit am Sourceende des Kanals stark von μ_{eff} ab, solange $\mu_{\text{eff}} \cdot E$ nicht sehr viel größer ist als die thermische Geschwindigkeit v_{T} . Das ist unter normalen Betriebsbedingungen ($V_{\text{DS}} \sim 1\text{V}$) selbst in extrem kurzen Kanälen ($L \sim 10\text{nm}$) der Fall, da zum einen die thermische Geschwindigkeit im On-Zustand exponentiell mit

der Elektronendichte zunimmt ($v_T > 2 \times 10^7 \text{ cm/s}$ [101]) und da zum anderen die Elektronenbeweglichkeit in MOSFET Inversionskanälen im Allgemeinen deutlich kleiner ist als der Maximalwert für undotiertes Bulk-Silizium, $\mu_{\text{max}} \approx 1500 \text{ cm}^2/\text{Vs}$ [36]. Im Grenzfall kleiner Feldstärken, also bei kleinen Drainspannungen oder in Transistoren mit langen Kanälen, folgt aus Gleichung (3-7) $v \approx \mu_{\text{eff}} \cdot E$. In diesem Fall ist der Drainstrom direkt proportional zur Beweglichkeit.

Der Elektronentransport in Silizium-Nanodrähten, d.h. in 1D Elektronensystemen, unterscheidet sich vom Transport in Bulk-Silizium (3D Elektronensystem) oder in gewöhnlichen MOSFET Kanälen (2D Elektronensysteme). Bisherige Untersuchungen zur Elektronenbeweglichkeit in SiNWs liefern widersprüchliche Ergebnisse. Während die Arbeiten von Kotlyar [136], Jin [137] und Ramayya [138] eine Abnahme der Beweglichkeit mit Verkleinerung des SiNW-Querschnitts zeigten, konnten Cui [139] und Koo [140] in SiNWs eine größere Elektronenbeweglichkeit als in konventionellen Bulk-MOSFETs feststellen. Diese Widersprüche lassen sich im Wesentlichen auf zwei gegenläufige Tendenzen zurückführen. Auf der einen Seite verringert sich beim Übergang von 2D zu 1D Systemen die Zustandsdichte für (gestreute) Elektronen, wodurch die Streuraten tendenziell abnehmen, was zu einer Zunahme der Elektronenbeweglichkeit führt [138]. Andererseits vergrößert sich bei Verkleinerung des Kanalquerschnitts der Überlappungsfaktor der Wellenfunktionen von Anfangs- und Endzustand der Elektronen bei Streuung durch Phononen, was tendenziell wiederum eine Verringerung der Beweglichkeit mit sich bringt [136], [141].

Im Folgenden werden Ergebnisse der theoretischen Untersuchungen zur Elektronenbeweglichkeit in SiNWs im Rahmen dieser Arbeit vorgestellt. Schwerpunkt der Untersuchungen war die Simulation der durch Phononenstreuung begrenzten Beweglichkeit in Tri-Gate MOS-Strukturen. Zunächst wird die Simulationemethode erläutert und im Anschluss daran werden die wichtigsten Ergebnisse diskutiert. Den Abschluss bildet eine kurze Zusammenfassung.

Simulationemethode

Die Methode zur Simulation der Elektronenbeweglichkeit in SiNWs basiert auf einer gekoppelten Lösung der 2D Schrödinger- und Poisson-Gleichungen einerseits sowie der 1D Boltzmann-Transportgleichung mit Hilfe von Monte-Carlo Simulationen andererseits. Im ersten Schritt werden mit NEXTNANO die 2D Schrödinger- und Poisson-Gleichungen für die untersuchten Tri-Gate MOS-Strukturen (Abb. 3.18)

selbstkonsistent gelöst. Die dabei berechneten Wellenfunktionen und Subbandenergien dienen als Eingangsgrößen für die 1D MC Transportsimulationen.

In den MC Simulationen wird ein konstantes elektrisches Feld in Transportrichtung angenommen, klein genug um Niedrigfeldtransport zu garantieren. Es werden verschiedene Streumechanismen berücksichtigt. Besonders wichtig bei kleinen senkrechten Feldern sind Streuprozesse an akustischen und optischen Phononen [141]. Streuung an akustischen Phononen lässt sich mit guter Näherung als elastischer Prozess modellieren [142]. Darüberhinaus ist die Annahme von Bulkphononen gerechtfertigt. Neueste Arbeiten welche die Quantisierung des Phononenspektrums berücksichtigen, zeigen einen Effekt auf die Beweglichkeit von höchstens 10% gegenüber der Annahme von Bulkphononen [138]. Unter Berücksichtigung dieser Annahmen ist die Streurrate für Elektronen mit dem Wellenvektor k_x an akustischen Phononen innerhalb eines Leitungsbandminimums gegeben durch [138]

$$\Gamma_{ij}^{\text{ac}}(k_x) = \frac{D_{\text{ac}}^2 k_B T \sqrt{2m^*}}{\hbar^2 \rho u^2} I_{ij} \frac{1 + 2\alpha \varepsilon_f}{\sqrt{\varepsilon_f(1 + \alpha \varepsilon_f)}} \Theta(\varepsilon_f). \quad (3-47)$$

Hier ist D_{ac} das Deformationspotential, m^* die Zustandsdichtemasse, ρ ist die Dichte von Silizium, u die Schallgeschwindigkeit in Silizium, $\alpha = 0,5$ ist der Nichtparabolizitätsfaktor, ε_f die kinetische Energie des gestreuten Elektrons und $\Theta(\varepsilon_f)$ ist die Heaviside-Sprungfunktion. Der Überlappungsfaktor I_{ij} beschreibt die räumliche Überlappung der Wellenfunktionen von Ausgangs- und Endzustand, $\psi_i(x,y)$ und $\psi_j(x,y)$, und ist gegeben durch [136]

$$I_{ij} = \iint |\psi_i(x,y)|^2 |\psi_j(x,y)|^2 dy dx, \quad (3-48)$$

wobei x und y die Ortskoordinaten in der Querschnittsebene der Tri-Gate MOS-Struktur sind.

Streuung an optischen Phononen ist verantwortlich für Elektronenübergänge zwischen verschiedenen Leitungsbandminima. Die entsprechenden Streuraten lassen sich mit

$$\Gamma_{ij}^{\text{iv}}(k_x) = \frac{D_{\text{iv}}^2 \sqrt{m^*}}{\sqrt{2\hbar \rho \omega_0}} \left(N_0 + \frac{1}{2} \mp \frac{1}{2} \right) I_{ij} \frac{1 + 2\alpha \varepsilon_f}{\sqrt{\varepsilon_f(1 + \alpha \varepsilon_f)}} \Theta(\varepsilon_f) \quad (3-49)$$

berechnen, wobei D_{iv} das Intervalley-Deformationspotential und $N_0 = [\exp(\hbar \omega_0 / k_B T) - 1]^{-1}$ die mittlere Anzahl von Phononen mit der Energie $\hbar \omega$ ist.

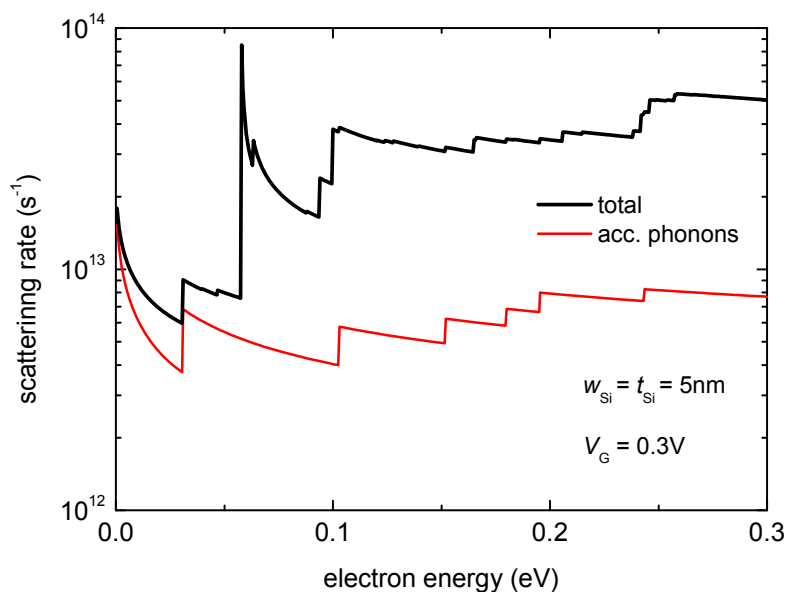


Abbildung 3.29 Streurrate von Elektronen im niedrigsten Subband an akustischen Phononen im Vergleich zur gesamten Streurrate für einen $5\text{nm} \times 5\text{nm}$ SiNW.

Wird ein Phonon während eines Streuvorgangs emittiert, erhöht sich die Phononenzahl um +1, bei Absorption bleibt die mittlere Anzahl unverändert N_0 .

Da die betrachteten SiNWs undotiert bzw. nur schwach p-dotiert sind, wurde Streuung an geladenen Störstellen nicht berücksichtigt.

Streuung an den Si/SiO₂-Grenzflächen wird bei großen senkrechten Feldern dominant [138]. Die Berücksichtigung von Grenzflächenstreuung im Rahmen von 1D Transportsimulationen ist wesentlich aufwendiger als die der bisher genannten Streumechanismen. Mit dem zur Verfügung stehenden 1D MC-Simulator kann Grenzflächenstreuung derzeit noch nicht beschrieben werden.

Abbildung 3.29 zeigt beispielhaft für einen $5\text{nm} \times 5\text{nm}$ SiNW die Streurrate von Elektronen im niedrigsten Subband an akustischen Phononen im Vergleich zur gesamten Streurrate, d.h. unter Berücksichtigung aller möglichen Elektronenübergänge zwischen den einzelnen Leitungsbandminima. Die Streuraten sind proportional zur elektronischen Zustandsdichte. Das wird in den sprunghaften Änderungen in der Streurrate deutlich. Diese markieren die Übergänge vom niedrigsten in ein energetisch höheres Subband. Entsprechend Gleichung (3-42) ist die Zustandsdichte für Elektronen an den Subbandkanten unendlich und fällt zu höheren Energien mit $\varepsilon_f^{-0.5}$ ab. In Abbildung 3.29 ist der Abfall jedoch aufgrund der hier berücksichtigten

Nichtparabolizität der Subbänder (Nichtparabolizitätsfaktor α) geringer. In Abbildung 3.29 wird außerdem deutlich, dass mit zunehmender kinetischer Energie der Elektronen die Wahrscheinlichkeit der Streuung an optischen Phononen (Intervalley-Streuung) im Vergleich zur Streuung an akustischen Phononen (Intravalley-Streuung) zunimmt.

Ergebnisse und Diskussion

Abbildung 3.30 zeigt simulierte Elektronenbeweglichkeiten für Tri-Gate MOS-Strukturen mit unterschiedlichen Siliziumquerschnitten als Funktion der Gatespannung. Alle Strukturen haben eine Gateoxiddicke von 1nm und die Austrittsarbeit der Gateelektrode ist gleich der Elektronenaffinität von Silizium. Deshalb liegt die Schwellspannung für jede Struktur bei ungefähr 0V. In den MC Simulationen wurde ein elektrisches Feld von 500 V/cm angenommen, das heißt es werden Niedrigfeldbeweglichkeiten berechnet.

In Abbildung 3.30 wird deutlich, dass die durch Phononenstreuung begrenzte Beweglichkeit mit Verkleinerung des Siliziumquerschnittes abnimmt. Das ist zunächst erstaunlich, da der energetische Abstand der Subbänder bei einer Verringerung des Querschnitts zunimmt, wodurch sich die Zustandsdichte für gestreute Elektronen verringert. Das wird jedoch mehr als kompensiert durch die sehr starke Zunahme des Überlappungsfaktors I_{ij} , was insgesamt zu einer Vergrößerung der Streuraten führt.

In Abbildung 3.31 sind die Überlappungsfaktoren für Streuung innerhalb des ersten Subbandes, I_{11} , als Funktion der Kanalabmessungen dargestellt. Die Symbole kennzeichnen I_{11} , die aus den mit NEXTNANO für $V_G = 0V$ simulierten Wellenfunktionen berechnet wurden, so wie sie tatsächlich zur Berechnung der Streuraten verwendet werden. Die Linie in Abbildung 3.31 zeigt I_{11} berechnet aus analytischen Wellenfunktionen, die man als Lösung der Schrödinger-Gleichung für das einfache 2D Potentialtopfmodell erhält. In diesem Fall liefert Gleichung (3-48) den Ausdruck $I_{11} = 9/(4w_{Si}t_{Si})$. Der Überlappungsfaktor ist also umgekehrt proportional zur Querschnittsfläche des Kanals. Die analytische Lösung für das Potentialtopfmodell stellt die Obergrenze für I_{11} dar, da aufgrund der angenommenen unendlich hohen Potentialbarriere die Wellenfunktionen auf den Siliziumkanal beschränkt sind. In NEXTNANO wird jedoch eine realistische endliche Barriere an der Si/SiO₂ Grenzfläche angenommen, so dass die Wellenfunktionen in das Oxid eindringen können und somit der effektive Kanalquerschnitt größer ist als $t_{Si} \times w_{Si}$. Die relative Abweichung von effektivem Kanalquerschnitt zum Siliziumquerschnitt nimmt mit

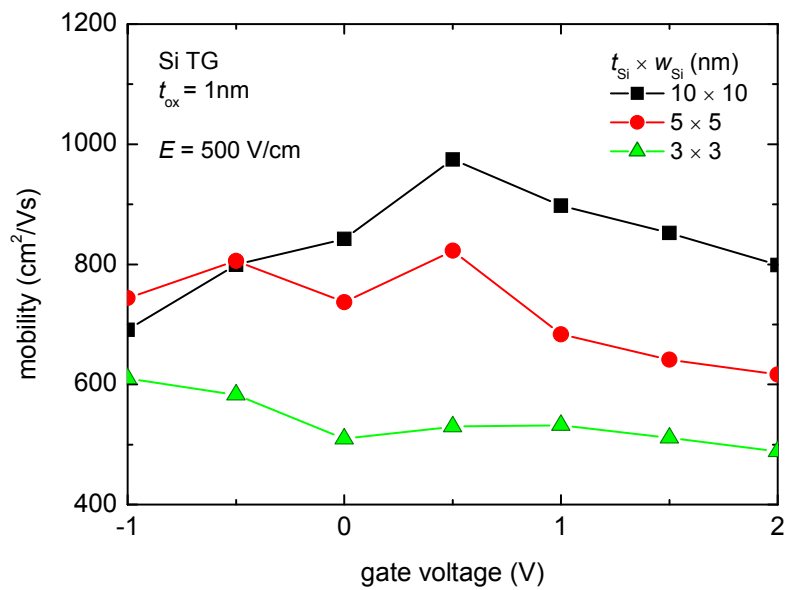


Abbildung 3.30 Durch Phononenstreuung begrenzte Elektronenbeweglichkeiten für Trigate MOS Strukturen mit verschiedenen Siliziumquerschnitten als Funktion der Gatespannung.

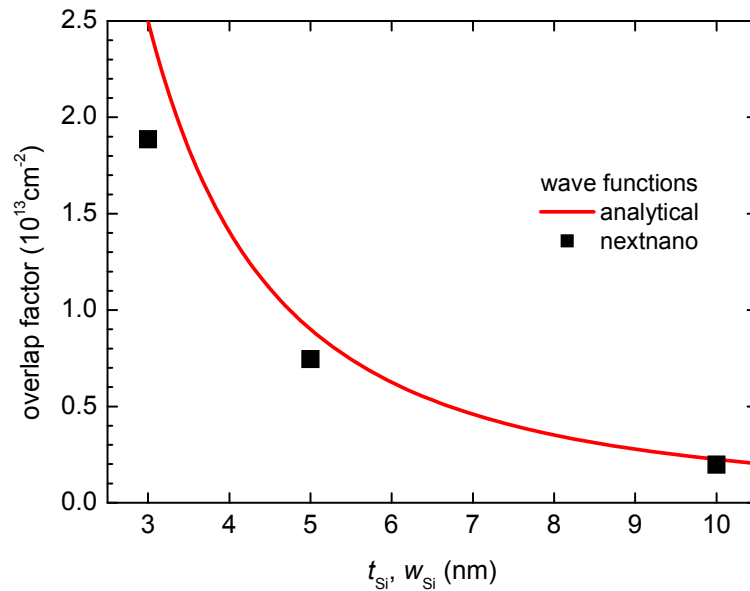


Abbildung 3.31 Überlappungsfaktoren für Streuung innerhalb des ersten Subbandes als Funktion der Kanal-abmessungen. Symbole kennzeichnen Überlappungsfaktoren berechnet aus mit NEXTNANO simulierten Wellenfunktionen. Die rote Linie kennzeichnet den Überlappungsfaktor für das 2D Potentialtopfmodell.

Verkleinerung der Siliziumabmessungen zu, so dass der Unterschied zwischen den Symbolen und der Linie in Abbildung 3.31 ebenfalls größer wird.

Die größte simulierte Beweglichkeit in Abbildung 3.30 beträgt $974\text{cm}^2/\text{Vs}$ für die $10\text{nm} \times 10\text{nm}$ Struktur. Das ist deutlich kleiner als die Elektronenbeweglichkeit in undotiertem Bulk-Silizium bei Raumtemperatur von etwa $1500\text{cm}^2/\text{Vs}$ [36]. Der Grund hierfür ist ebenfalls in dem sich vergrößernden Überlappungsfaktor zu suchen. Die hier berechneten Beweglichkeiten sind vergleichbar mit experimentellen Werten für SiNWs mit ähnlichem Querschnitt [140].

In Abbildung 3.32 ist die für eine $5\text{nm} \times 5\text{nm}$ Tri-Gate-Struktur simulierte Beweglichkeit als Funktion der Linienkonzentration der Elektronen dargestellt. Die Beweglichkeit hat ein ausgeprägtes Maximum bei einer Liniendichte von 10^7cm^{-1} ($V_G = 0,5\text{V}$). Dieses Maximum hat seine Ursache im Wechselspiel zweier gegenläufiger Tendenzen. Mit zunehmender Elektronendichte wird die Bandverbiegung im Silizium stärker, wodurch die Elektronen näher an die Si/SiO₂ Grenzflächen gezogen werden. Das ist gleichbedeutend mit einer Verringerung des effektiven Kanalquerschnitts. Folglich vergrößern sich die Überlappungsfaktoren für die verschiedenen Streuprozesse, was tendenziell zu einer Verringerung der Beweglichkeit führt. Auf der anderen Seite wird der Potentialverlauf im Silizium zunehmend unsymmetrischer wodurch es zu einer immer stärkeren Verschiebung der Subbänder verschiedener Leitungsbandminima gegeneinander kommt. Dadurch wird Intervalley-Streuung zunehmend unwahrscheinlicher, was die Beweglichkeit tendenziell erhöht. Beide Trends werden in Abbildung 3.33 offensichtlich. Sowohl der Überlappungsfaktor I_{11} als auch der energetische Abstand der beiden niedrigsten Subbänder ΔE_{12} (die beide zu verschiedenen Minima gehören) werden mit zunehmender Elektronenkonzentration größer. Jedoch ändert sich I_{11} bei Konzentrationen zwischen 10^6 und 10^7cm^{-1} im Gegensatz zu ΔE_{12} nur geringfügig, was den Anstieg der Beweglichkeit bis zum Maximalwert erklärt.

Fazit

Die Elektronenbeweglichkeit in Silizium-Nanodrähten wird durch Effekte des Quantum Confinement beeinflusst. Die Simulationen ergaben, dass eine Verkleinerung des Siliziumquerschnitts prinzipiell zu einer Verringerung der Beweglichkeit führt. Das ist auf die sehr starke Zunahme des Überlappungsfaktors zurückzuführen. Eine noch stärkere Abhängigkeit der Elektronenbeweglichkeit vom Kanalquerschnitt ist bei einer Berücksichtigung von Grenzflächenstreuung zu erwarten. Eine Berücksichtigung von

Grenzflächenstreuung war jedoch im Rahmen dieser Arbeit nicht möglich. Die simulierten Beweglichkeiten für die größten hier betrachteten Strukturen sind vergleichbar mit aktuellen experimentellen Ergebnissen.

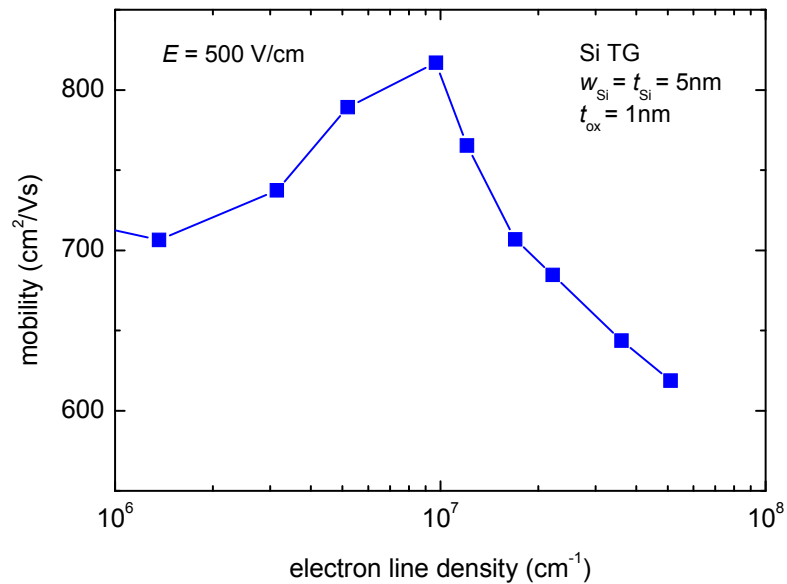


Abbildung 3.32 Simulierte Elektronenbeweglichkeit für eine $5\text{nm} \times 5\text{nm}$ SiNW-Struktur als Funktion der Linienkonzentration der Elektronen.

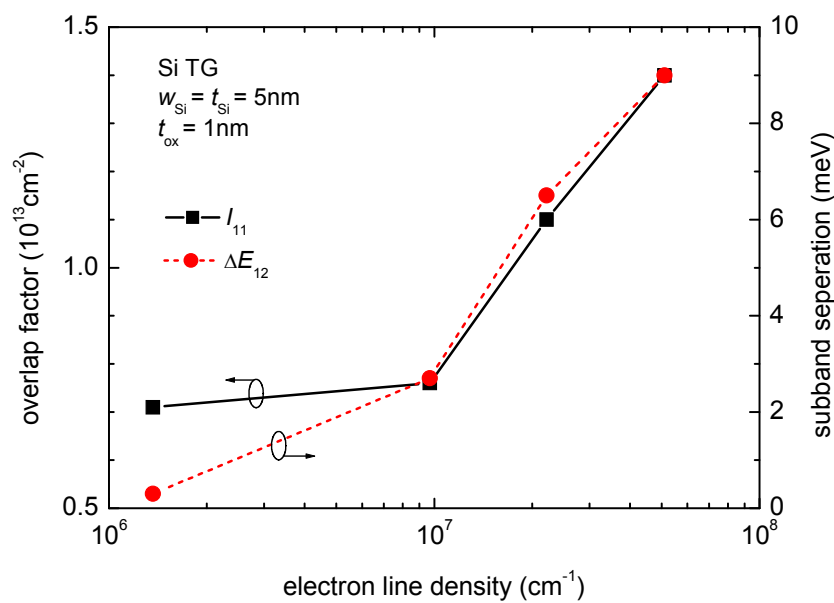


Abbildung 3.33 Überlappungsfaktor und energetischer Abstand der beiden niedrigsten Subbänder als Funktion der Linienkonzentration der Elektronen.

3.4.4. Korrektur klassischer Bauelementesimulationen

Die in dieser Arbeit untersuchten Quanteneffekte in Nanometer-MOSFETs bewirken makroskopisch eine Abweichung von klassisch berechneten Kennlinien. Nachfolgend werden einfache Möglichkeiten besprochen, die Wirkung der Quanteneffekte durch Korrekturen klassischer Bauelementesimulationen bzw. der damit simulierten Kennlinien zu berücksichtigen.

Verschiebung der Schwellspannung

Die Berücksichtigung der durch die Subbandformierung verursachten Schwellspannungs-verschiebung wurde bereits in Abschnitt 3.4.1. diskutiert. Ist für einen Transistor die klassisch simulierte Schwellspannung und der Subthreshold Slope vorhanden, lässt sich die tatsächliche Schwellspannung ganz einfach mit Hilfe von (3-33) und (3-34) berechnen. Voraussetzung ist ein Modell für die Verschiebung des Kanalpotentials im betrachteten MOSFET. Gleichung (3-35) stellt ein adäquates Modell für vollständig verarmte DG und SG SOI MOSFETs dar und Gleichung (3-31) ein Modell für MG MOSFETs mit rechteckigem Querschnitt.

Wie sich mit Hilfe dieser Modelle die Auswirkung von Quantum Confinement auf Schwankungen von I_{on} und I_{off} infolge von Geometrietoleranzen berücksichtigen lassen, wird in Abschnitt 4.2.4. ausführlich behandelt.

Gatekapazität und Steilheit

Die durch Quanteneffekte verringerte der Gatekapazität macht sich gleichstrommäßig als geringere Steilheit bemerkbar. Dadurch wird im Wesentlichen der On-Strom reduziert.

Zur Illustration dieses Effekts sind in Abbildung 3.34 mit SCHRED quantenmechanisch sowie klassisch simulierte n_s in DG MOS-Strukturen als Funktion der effektiven Gatespannung $V_{G,\text{eff}} = V_G - V_{\text{th}}$ zu sehen. Im Subthresholdbereich (Abbildung 3.34a) liegen die klassischen und quantenmechanischen Kennlinien übereinander, d.h. hier spielt nur die Verschiebung der Schwellspannung eine Rolle. Dagegen wird in der linearen Darstellung in Abbildung 3.34(b) deutlich, dass für jede Struktur ab $V_G - V_{\text{th}} > 0$ die quantenmechanische Elektronenschichtkonzentration wesentlich kleiner ist als die aus klassischen Simulationen. Dieser Effekt ist umso größer, je kleiner die äquivalente Oxiddicke EOT ist, was sich mit der Reihenschaltung von C_{ox} und C_{inv} [vgl. Abb. 3.17 und Gleichung (3-37)] erklären lässt. Entsprechend Gleichung (3-18) wirkt sich die

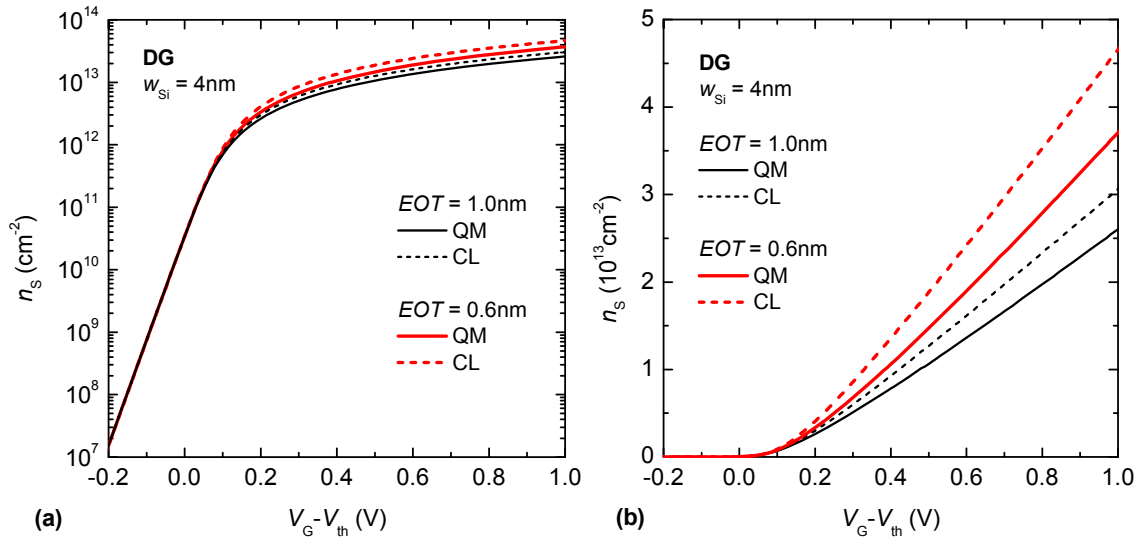


Abbildung 3.34 Mit SCHRED klassisch und quantenmechanisch berechnete Elektronenschichtkonzentration als Funktion der effektiven Gatespannung für DG MOS-Strukturen mit verschiedenen EOT . (a) Logarithmischer und (b) linearer Maßstab.

verringerte Elektronenschichtkonzentration direkt auf den Drainstrom eines MOSFET aus.

Um die Auswirkungen des o.g. Effekts für die in dieser Arbeit relevanten MOSFET-Designs abzuschätzen, wurde für eine Reihe verschiedener SG und DG MOS-Strukturen das Verhältnis von quantenmechanischer und klassischer Elektronenschichtkonzentration n_s^{QM}/n_s^{CL} mit SCHRED bestimmt. Abbildung 3.35 zeigt n_s^{QM}/n_s^{CL} als Funktion der effektiven Gatespannung für DG-Strukturen mit $EOT = 0.6 \text{ nm}$. Darin wird deutlich, dass das Verhältnis n_s^{QM}/n_s^{CL} stark von der Gatespannung und von der Si-Dicke abhängt. Um die Auswirkung des Quanteneffekts auf den On-Strom eines MOSFET abzuschätzen, muss demnach n_s^{QM}/n_s^{CL} für das entsprechende $V_{G,eff}$ extrahiert werden. Für dieses $V_{G,eff}$ lässt sich dann sehr leicht aus dem klassisch berechneten Drainstrom I_D^{CL} der tatsächliche Drainstrom bestimmen.

$$I_D(V_{G,eff}) \approx \frac{n_s^{QM}}{n_s^{CL}}(V_{G,eff}) \cdot I_D^{CL}(V_{G,eff}) \quad (3-50)$$

Abbildung 3.36 zeigt n_s^{QM}/n_s^{CL} für $V_{G,eff} = 0.6 \text{ V}$ als Funktion der Siliziumdicke für SG und DG MOS-Strukturen mit unterschiedlichen EOT . Es wird deutlich, dass sich

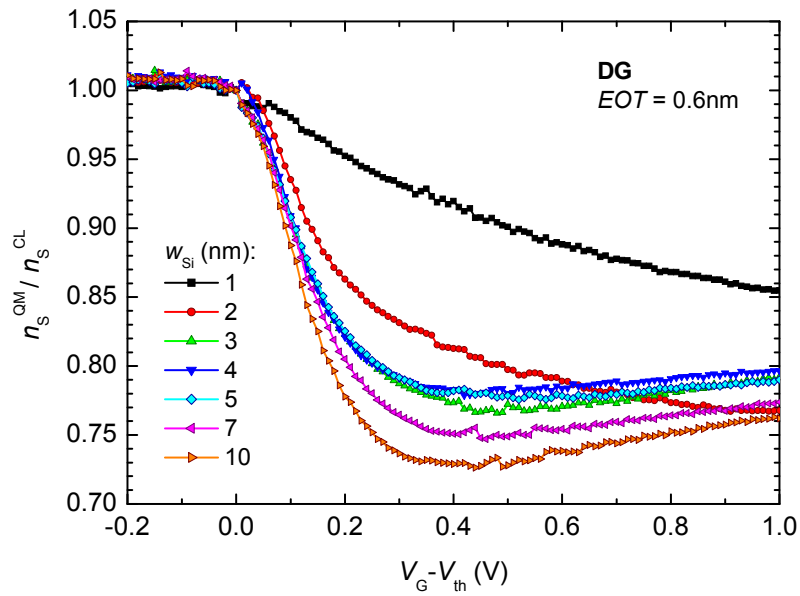


Abbildung 3.35 Verhältnis von quantenmechanischer zu klassischer Elektronenschichtkonzentration als Funktion der effektiven Gatespannung für DG MOS-Strukturen mit verschiedenen Schichtdicken.

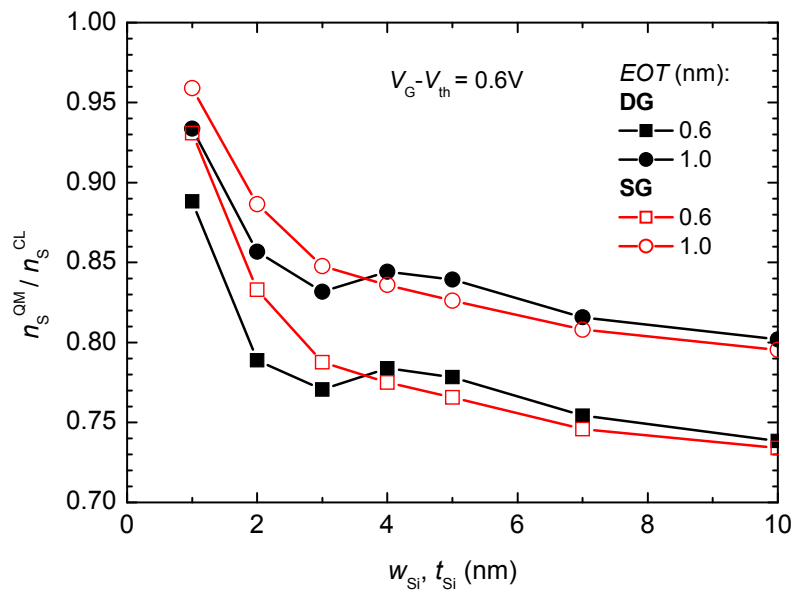


Abbildung 3.36 Verhältnis von quantenmechanischer zu klassischer Elektronenschichtkonzentration bei $V_G - V_{th} = 0.6V$ als Funktion der Siliziumdicke für SG und DG Strukturen mit verschiedenen EOT.

$n_s^{\text{QM}}/n_s^{\text{CL}}$ mit abnehmender Si-Dicke dem Wert 1 annähert. Das liegt daran, dass sowohl klassisch als auch quantenmechanisch der Ladungsschwerpunkt x_{av} umso näher an der Si-Oberfläche liegen muss, je dünner die Si-Schicht ist. Mit abnehmendem x_{av} wird aber die elektrostatische Kapazität des Kanals $C_{\text{es,Si}} = \epsilon_{\text{Si}} / x_{\text{av}}$ größer und deren Einfluss auf C_G entsprechend Gleichung (3-37) geringer. Folglich machen sich die Unterschiede zwischen quantenmechanisch und klassisch berechneten Elektronenverteilungen im Kanal immer weniger in unterschiedlichen n_s bemerkbar.

Generell sind sich in Abbildung 3.36 die $n_s^{\text{QM}}/n_s^{\text{CL}}$ für SG und DG Strukturen mit gleichem EOT bei allen Schichtdicken sehr ähnlich. Für Si-Dicken kleiner als etwa 3nm sind die Unterschiede etwas größer. In diesem Bereich liegt $n_s^{\text{QM}}/n_s^{\text{CL}}$ für SG-Strukturen näher an 1 als für DG-Strukturen. Der Grund dafür ist, dass in SG SOI-Strukturen die Elektronendichte an der Grenzfläche zum BOX sowohl klassisch als auch quantenmechanisch relativ klein ist, so dass Unterschiede keinen so großen Einfluss auf die integrale Elektronendichte n_s haben.

Die Verringerung von Gatekapazität und Steilheit wirkt sich auch auf das Kleinsignalverhalten extrem skalierten MOSFETs aus. Die Kleinsignalparameter, in welche die Größen g_m , C_{GS} oder C_{GD} eingehen (siehe Anhang A), hängen von der Stärke dieses Effekts ab. Da aber entsprechend Gleichung (3-13) die Verringerung von g_m im Zähler durch die von C_{GG} im Nenner kompensiert wird, sollte sich die Quantisierung nicht wesentlich auf f_T auswirken. Diese Überlegung wird durch die Ergebnisse eines direkten Vergleiches von quantenmechanischen und klassischen Simulationen eines DG MOSFET mit $L_G = 15\text{nm}$ [143] bestätigt. Im Rahmen dieser Arbeit wird deshalb auf die Berücksichtigung von Quanteneffekten bei der Simulation von Hochfrequenzeigenschaften verzichtet.

Elektronenbeweglichkeit

Die effektive Beweglichkeit in 1D Elektronensystemen lässt sich prinzipiell über ein geeignetes Modell in jedem klassischen Bauelementesimulator berücksichtigen. Wie wir in Abschnitt 3.4.3. gesehen haben muss ein solches Modell neben der Wirkung verschiedener Streumechanismen den Einfluss der Kanalgeometrie richtig beschreiben. Leider gibt es ein solches Modell bisher noch nicht. Im Rahmen dieser Arbeit wurde deshalb auch für die Simulation von MG MOSFETs das herkömmliche Lombardi-Modell für planare MOSFETs [94] verwendet.

Kapitel 4

Nanometer-MOSFETs für digitale Logik

4.1. Einführung

Die digitale Logik ist die Grundlage der modernen Informationstechnologie. In modernen elektronischen Systemen gewinnt sie immer mehr an Bedeutung, da analoge Funktionen zunehmend durch digitale Schaltungen ersetzt werden [144]. Die überwiegende Mehrzahl aller digitalen Schaltungen wird in der so genannten CMOS Logik (complementary MOS) realisiert. Die Grundgatter dieser Logikfamilie sind aus komplementären MOSFETs, das heißt aus n- und pMOSFETs aufgebaut. Das einfachste Logikgatter, der CMOS-Inverter (Abb. 4.1), besteht aus einem n-Kanal und einem p-Kanal MOSFET. Dieses Konzept hat den Vorteil, dass im stationären Zustand, d.h. High am Eingang / Low am Ausgang oder umgekehrt, grundsätzlich einer der beiden Transistoren gesperrt ist. Dadurch ist die statische Verlustleistung in CMOS-Schaltungen im Vergleich zu anderen Logikfamilien sehr gering. Die geringe Verlustleistung sowie der relativ einfache technische Aufbau sind wesentliche Gründe für den Erfolg von CMOS und somit des Silizium-MOSFET als Grundbauelement in der Digitaltechnik.

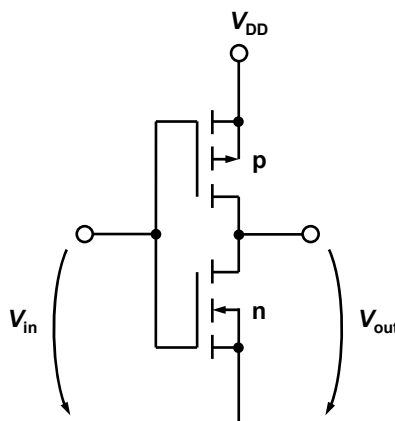


Abbildung 4.1 CMOS Inverter.

MOSFET-Konzepte, die auf der Verwendung von Siliziumnanodrähten basieren, werden oft als die ultimativen Logikbauelemente für eine Skalierung in den Gatelängenbereich um 10nm angesehen [4], [63]. Der Grund dafür ist die exzellente Kontrolle des Kanalpotentials durch das Gate. Beim Tri-Gate Konzept (siehe Abb. 4.2), beispielsweise, ist das Gate von drei Seiten elektrostatisch an den Kanal gekoppelt. Bei diesem Transistorkonzept hat man im Wesentlichen zwei Freiheitsgrade hinsichtlich der Querschnittsgeometrie, nämlich die Höhe t_{Si} und die Weite w_{Si} . Es stellt sich nun die Frage nach dem optimalen Kanaldesign, d.h. nach der Wahl geeigneter Querschnitte für MOSFETs mit einer Gatelänge von nur 10nm.

Die hier vorgestellten Untersuchungen konzentrieren sich auf das Design von 3D Tri-Gate (TG) MOSFETs mit einer Gatelänge von 10nm sowie der beiden 2D Grenzfälle Single-Gate (SG) SOI MOSFET und Double-Gate (DG) MOSFET. Die Massenproduktion von MOSFETs mit einer Gatelänge von 10nm wird um das Jahr 2020 gefordert [4]. Als Auswahlkriterium für geeignete Designs werden hier die für Digital-MOSFETs wichtigen Parameter On-Strom I_{on} und Off-Strom I_{off} verwendet. Das bedeutet, der Drainstrom des Transistors darf im ausgeschalteten Zustand höchstens $I_{\text{off}}^{\text{max}}$ sein und muss im eingeschalteten Zustand mindestens $I_{\text{on}}^{\text{min}}$ erreichen.

In den folgenden Abschnitten wird zunächst die Simulationemethode erläutert. Danach werden die Ergebnisse zum Design des inneren Transistors diskutiert. Die Auswirkung von Geometrieschwankungen auf die Parameter V_{th} , I_{on} und I_{off} wird analysiert und der Effekt von parasitären Widerständen untersucht.

4.2. Design von Single-Gate, Double-Gate und Tri-Gate MOSFETs

4.2.1. Simulationemethode

Die Simulationsstrukturen sind in Abbildung 4.2 skizziert. Tri-Gate MOSFETs besitzen eine dreidimensionale Struktur und erfordern daher 3D Simulationen. Double-Gate und Single-Gate MOSFETs stellen die zweidimensionalen Grenzfälle der TG-Struktur dar und können mit 2D Simulationen beschrieben werden. Den DG MOSFET erhält man aus der TG-Struktur für $t_{\text{Si}} \rightarrow \infty$, den SG MOSFET für $w_{\text{Si}} \rightarrow \infty$. Aufgrund der symmetrischen Struktur von TG und DG MOSFETs genügt für deren vollständige

Beschreibung jeweils die Simulation der halben Struktur. Das führt besonders in den sehr aufwendigen 3D Simulationen zu einer erheblichen Reduzierung der Rechenzeiten.

Alle Strukturen besitzen einen undotiertem Kanal, high-K Dielektrika und Metall-Gates. Die Abmessungen und Dotierungen der Simulationsstrukturen sind in Tabelle 4.1 zusammengefasst. Um eine äquivalente Gateoxiddicke EOT von 0.6nm einzustellen wurde ein idealer Gateisolator mit $\epsilon_r = 6.5$ angenommen¹. Die Dicke des BOX entspricht den Vorgaben der ITRS ($t_{BOX} \approx 2L_G$). Das Rückseitensubstrat der SG- und TG-Strukturen wurde vereinfacht als Metallelektrode auf Massepotential (0V) modelliert, um den numerischen Aufwand der 3D-Simulationen zu begrenzen. Vergleichssimulationen haben gezeigt, dass eine realistischere Modellierung der Rückseite nahezu die gleichen Ergebnisse liefert.

Die Source- und Draingebiete sind mit einer Länge von 5nm sehr kurz und die Übergänge zum Kanal sind abrupt. Dieses idealisierte Design wurde gewählt, damit die Serienwiderstände zunächst möglichst klein sind. Der Einfluss von Serienwiderständen wird dann in Abschnitt 4.3 gesondert betrachtet.

Alle Simulationen wurden mit dem Bauelementesimulator ATLAS unter Verwendung des Drift-Diffusionsmodells durchgeführt. Effekte des nichtstationären Ladungstransports wurden mit der in Abschnitt 3.2.2. vorgestellten modifizierten $v(E)$ -Charakteristik [92] berücksichtigt. Da bisher noch kein Modell für die Niedrigfeldbeweglichkeit in SiNW MOSFETs existiert, wurde das Lombardi-Modell [94] verwendet, das die Beweglichkeit in Bulk-MOSFETs beschreibt. Es wurde mit Fermi-Dirac-Statistik gerechnet. Quanteneffekte wurden nicht direkt während der Simulationen sondern im Sinne von nachträglichen Korrekturen der zunächst klassischen Ergebnisse berücksichtigt. Der Einfluss von Quantum Confinement auf die Schwellspannung oder den Off-Strom wurde wenn nötig mit Hilfe analytischer Modelle [66] berücksichtigt. Für die Korrektur der simulierten klassischen On-Ströme wurden Korrekturfaktoren verwendet, welche mit Hilfe von Schrödinger-Poisson Simulationen bestimmt wurden.

¹ Die Kernaussagen dieser Untersuchungen gelten grundsätzlich für alle Gateisolatoren mit $EOT = 0.6\text{nm}$, inklusive Hafnium-basierter Dielektrika mit wesentlich größeren ϵ_r .

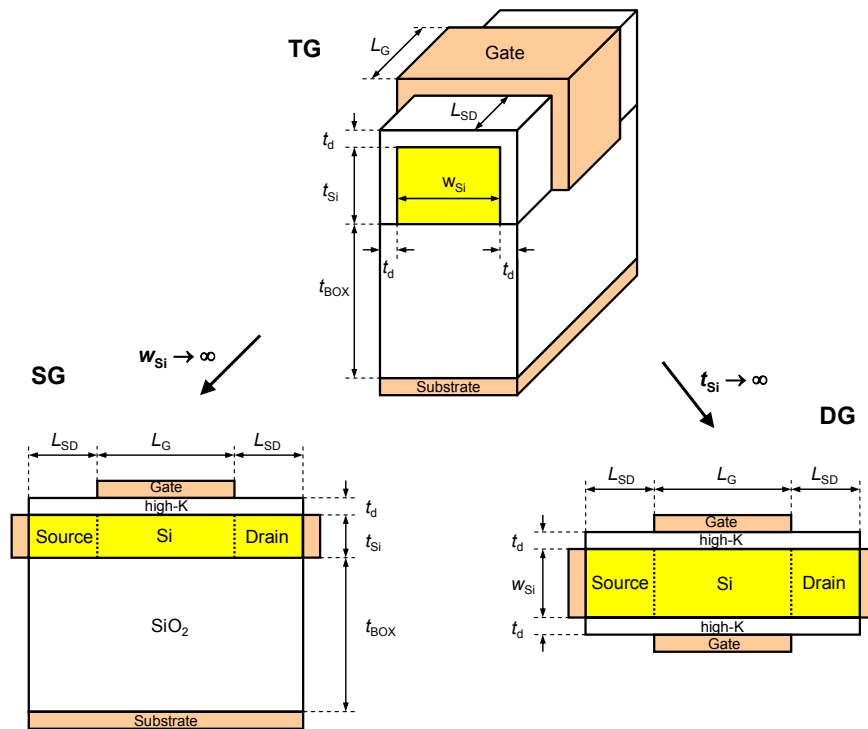


Abbildung 4.2 Die Simulationsstrukturen für Single-Gate, Double-Gate und Tri-Gate MOSFETs. Die Abmessungen und Dotierungen sind in Tabelle 4.1 zusammengefasst.

Gatelänge, L_G (nm)	10
Siliziumdicke, t_{Si} (nm)	variabel
Siliziumweite, w_{Si} (nm)	variabel
Länge der Source- und Draingebiete, L_{SD} (nm)	5
Äquivalente Oxiddicke, EOT (nm)	0.6
Dicke des Gateisolators, t_d (nm)	1
Relative DEK des Gatedisolators, ϵ_r	6.5
Dicke des vergrabenen Oxids (SG, TG), t_{BOX} (nm)	20
Source/Drain-Dotierung, N_D (cm ⁻³)	2×10^{20}
Kanaldotierung, N_A (cm ⁻³)	10^{10}
Gateaustrittsarbeit, Φ_G (eV)	variabel
Austrittsarbeit des Substratkontakts (SG, TG), Φ_M (eV)	4.72
Betriebsspannung, V_{DD} (V)	0.8

Tabelle 4.1 Designparameter der in Abbildung 4.2 skizzierten Simulationsstrukturen.

4.2.2. Single-Gate und Double-Gate MOSFETs

Betrachten wir zunächst den zweidimensionalen Grenzfall DG MOSFET. In Abbildung 4.3 sind Transferkennlinien ohne Quantenkorrekturen für DG MOSFETs mit verschiedenen w_{Si} für eine Drain-Source Spannung von $V_{\text{DS}} = V_{\text{DD}} = 0.8\text{V}$ zu sehen. In den Simulationen wurde für alle Strukturen eine Gateaustrittsarbeit von 4.72eV verwendet, d.h. es wurde Midgap-Material angenommen. Der Drainstrom ist in Abbildung 4.3 auf die effektive Gateweite w_{eff} normiert. Entsprechend Abbildung 4.2 gilt für den DG MOSFET $w_{\text{eff}} = 2t_{\text{Si}}$, für den SG MOSFET $w_{\text{eff}} = w_{\text{Si}}$ und für den TG MOSFET $w_{\text{eff}} = w_{\text{Si}} + 2t_{\text{Si}}$.

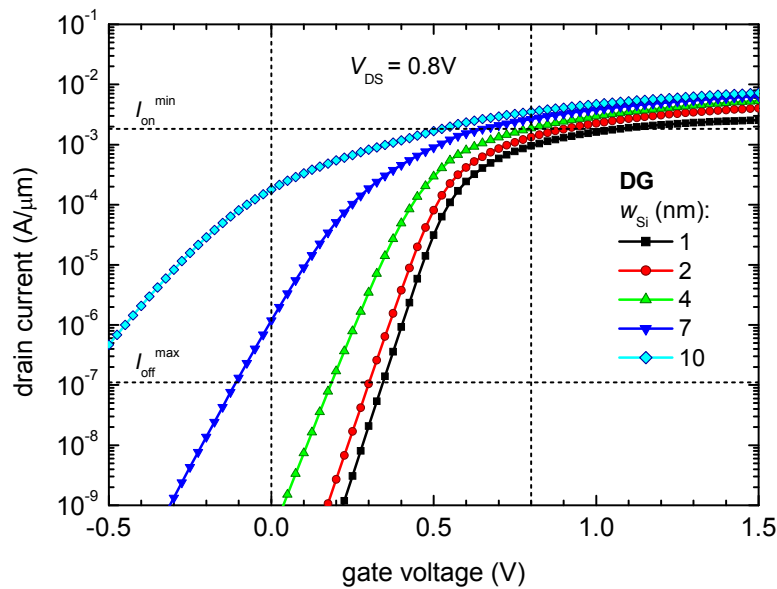


Abbildung 4.3 Simulierte Transferkennlinien von DG MOSFETs mit verschiedenen w_{Si} ohne Quantenkorrekturen. Horizontale Linien markieren die Zielgrößen für I_{on} und I_{off} . Die vertikalen Linien begrenzen den Gatespannungsbereich für digitale Anwendungen bei der hier betrachteten Betriebsspannung von 0.8V .

In Abbildung 4.3 wird deutlich, dass sich das Abschaltverhalten verbessert, je kleiner w_{Si} gemacht wird. Der Subthreshold Slope verringert sich mit abnehmendem w_{Si} aufgrund der stärkeren elektrostatischen Kopplung des Gates an den Kanal. Dadurch kann die Höhe der Potentialbarriere zwischen Source und Drain besser durch die Gatespannung gesteuert werden. Gleichzeitig verschieben sich die Kennlinien in Richtung positive V_{GS} . Diese Verschiebung hat im Wesentlichen zwei Ursachen. Zum einen sind Kurzkanaleffekte, wie DIBL oder V_{th} -Roll-Off, bei kleineren w_{Si} deutlich

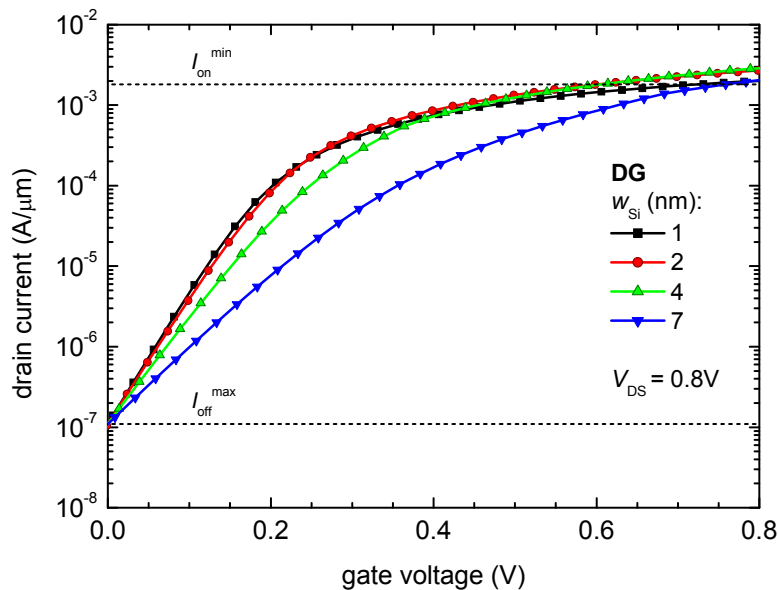


Abbildung 4.4 Transferkennlinien der DG MOSFETs aus Abb. 4.3 so verschoben, dass bei $V_{GS} = 0V$ das I_{off} -Kriterium gerade erfüllt ist.

geringer. Andererseits verringert sich mit abnehmendem Kanalquerschnitt die Elektronenschichtkonzentration für ein bestimmtes V_{GS} , so dass I_D kleiner wird.

Die horizontalen Linien in Abbildung 4.3 markieren die beiden Zielgrößen I_{off}^{max} und I_{on}^{min} , so wie sie typischerweise für High-Performance Logik festgelegt werden [4]. Die vertikalen Linien begrenzen den Gatespannungsbereich für digitale Anwendungen bei der hier betrachteten Betriebsspannung von 0.8V. Von den gezeigten Kennlinien erfüllt nur die des DG MOSFET mit $w_{Si} = 4nm$ sowohl das Kriterien für I_{on} als auch das Kriterium für I_{off} innerhalb dieses Gatespannungsintervalls. Die Lage der Kennlinie auf der Gatespannungsachse ist jedoch nicht optimal. Theoretisch ist es möglich, über die Wahl des Gatematerials (d.h. über Φ_G) die Transferkennlinie auf der V_{GS} -Achse beliebig zu verschieben. Abbildung 4.4 zeigt die Transferkennlinien der Transistoren aus Abbildung 4.3 so verschoben, dass alle das I_{off} -Kriterium bei $V_{GS} = 0V$ gerade erfüllen. Der Drainstrom der nun bei $V_{GS} = V_{DS} = 0.8V$ fließt ist der maximal mögliche On-Strom eines bestimmten MOSFET-Designs bei gleichzeitiger Erfüllung des I_{off} -Kriteriums.

In Abbildung 4.5 sind die so ermittelten I_{on} für DG MOSFETs als Funktion von w_{Si} und für SG MOSFETs als Funktion von t_{Si} dargestellt. Neben den klassisch simulierten On-Strömen I_{on}^{CL} sind auch korrigierte Werte gezeigt, welche die Wirkung von Quanteneffekten berücksichtigen. Die korrigierten On-Ströme wurden durch

Multiplikation mit einem Korrekturfaktor bestimmt, $I_{\text{on}} = c^{\text{QM}} \cdot I_{\text{on}}^{\text{CL}}$. Der Korrekturfaktor c^{QM} gibt das Verhältnis von quantenmechanischer und klassischer Elektronenschichtkonzentration an, welches für jede Struktur mit Hilfe von Schrödinger-Poisson Simulationen [117] bestimmt wurde (vgl. Abb. 3.36). Um den Rechenaufwand zu minimieren wurde darüber hinaus eine vereinfachte Quantenkorrektur mit einem konstanten Korrekturfaktor für jede Struktur getestet. Es zeigt sich, dass ein konstanter Korrekturfaktor von 0.8 ähnliche Ergebnisse liefert, und zwar nicht nur für DG und SG MOSFETs sondern auch für TG MOSFETs.

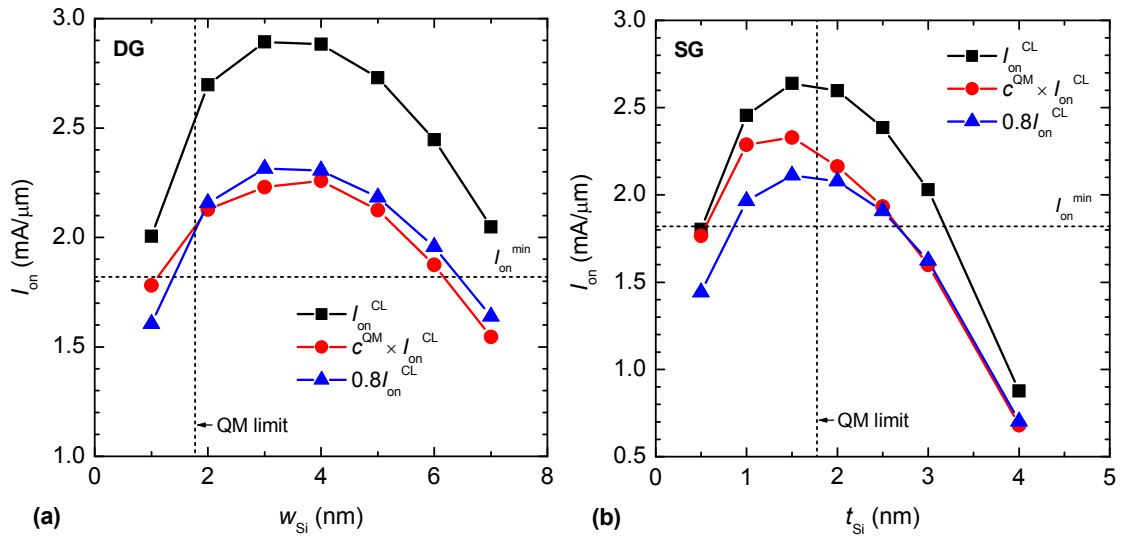


Abbildung 4.5 Maximal möglicher On-Strom als Funktion der Silizium-Schichtdicke für (a) DG und (b) SG MOSFETs bei gleichzeitiger Erfüllung des I_{off} -Kriteriums. Gezeigt sind die On-Ströme aus der klassischen Simulation, sowie zwei Werte mit einer Quantenkorrektur: variabler Korrekturfaktor $c^{\text{QM}}(w_{\text{Si}}, t_{\text{Si}})$, konstanter Korrekturfaktor 0.8 für alle Strukturen. Gestrichelte Linien – horizontal: Target für I_{on} , vertikal: durch Quanteneffekte definierte minimale Schichtdicke.

Der Bereich von w_{Si} (t_{Si}) in dem das I_{on} -Kriterium erfüllt ist, hat eine obere und eine untere Grenze, $w_{\text{Si,max}}$ ($t_{\text{Si,max}}$) und $w_{\text{Si,min}}$ ($t_{\text{Si,min}}$). Bestimmend für die obere Grenze ist das Abschaltverhalten. Wenn w_{Si} kleiner wird, kann auch mit einer geringeren Schwellspannung [vgl. Definition (2-1)] das I_{off} -Kriterium erfüllt werden. Dadurch wird die effektive Gatespannung $V_{\text{GS}} - V_{\text{th}}$ im On-Zustand größer, wodurch I_{on} steigt.

Die untere Grenze des Bereiches von w_{Si} bzw. t_{Si} in dem das I_{on} -Kriterium erfüllt ist, wird von den Widerständen der Source- und Draingebiete, R_{S} und R_{D} , bestimmt. Diese lassen sich für die Simulationsstrukturen folgendermaßen abschätzen

$$R_S = R_D = \frac{1}{qN_D\mu_n} \cdot \frac{L_{SD}}{t_{Si}w_{Si}} \cdot w_{eff}, \quad (4-1)$$

wobei R_S und R_D auf die jeweilige effektive Gateweite normiert sind. Bei Verringerung des Kanalquerschnitts vergrößern sich R_S und R_D und I_{on} wird kleiner. Da beim DG MOSFET w_{eff} doppelt so groß ist wie beim SG MOSFET, sind auch R_S und R_D bei gleicher Schichtdicke doppelt so groß. Aus diesem Grund ist in Abbildung 4.5 die minimale Schichtdicke zur Erfüllung des I_{on} -Kriteriums beim DG MOSFET, $w_{Si,min}$, deutlich größer als $t_{Si,min}$ beim SG SOI MOSFET.

Bei Siliziumdicken von wenigen Nanometern sind starke Quanteneffekte zu erwarten. Diese bewirken, dass Geometriefluktuationen zu erheblichen Unterschieden von V_{th} bzw. I_{off} zwischen nominell gleichen Transistoren führen können. Es ist daher sinnvoll, eine quantenmechanisch definierte untere Grenze für die Schichtdicke, $w_{Si,min}^{QM}$ bzw. $t_{Si,min}^{QM}$, einzuführen. Eine vernünftige Forderung wäre beispielsweise, dass bei Schwankungen der Querschnittsgeometrie um maximal eine Atomlage ($\Delta w_{Si} = \Delta t_{Si} = 1AL \approx 1.36\text{\AA}$) das Verhältnis vom tatsächlichen zum nominellen Off-Strom nicht größer als 2 sein darf. Für DG und SG MOSFETs mit Langkanaleigenschaften lassen sich die minimalen Schichtdicken, für die diese Forderung gerade noch erfüllt ist, mit Hilfe der Funktion β aus Gleichung (3-35) bestimmen.

$$\frac{I_{off}}{I_{off}^{max}} \approx \frac{n_S^{QM}(w_{Si,min}^{QM} + \Delta w_{Si})}{n_S^{QM}(w_{Si,min}^{QM})} = \frac{\beta(w_{Si,min}^{QM} + \Delta w_{Si})}{\beta(w_{Si,min}^{QM})} = 2 \quad (4-2)$$

Man erhält als minimale Abmessung $w_{Si,min}^{QM} = 1.78\text{nm}$. Da im Subthresholdbereich die Stärke der Quanteneffekte nur von der Si-Schichtdicke abhängt, sind die quantenmechanisch definierten minimalen Schichtdicken für DG und SG MOSFETs gleich groß, $w_{Si,min}^{QM} = t_{Si,min}^{QM}$. Sie sind außerdem größer als die mit dem I_{on} -Kriterium bestimmten minimalen Schichtdicken $w_{Si,min}$ und $t_{Si,min}$, so dass der Bereich nutzbarer Schichtdicken in Abbildung 4.5 beim SG-Design wesentlich stärker durch Quanteneffekte eingeschränkt wird als beim DG-Design. Für den SG SOI MOSFET verbleibt ein Designspielraum für t_{Si} von weniger als 1nm, für den DG MOSFET ein Spielraum für w_{Si} von mehr als 4nm. Mit anderen Worten, das DG-Konzept gestattet dem Designer eine größere Freiheit bezüglich der Schichtdicke als das SG-Konzept.

4.2.3. Tri-Gate MOSFETs

Für die Gleichstromeigenschaften von TG MOSFETs mit $L_G = 10\text{nm}$ (Abb. 4.2) ist die Gestaltung der Querschnittsgeometrie entscheidend. Die Frage dabei ist, welche Kombinationen von w_{Si} und t_{Si} das Einhalten der Kriterien für I_{on} und I_{off} ermöglichen.

Um den Designspielraum einzugrenzen, wurde für eine Anzahl von TG Designs, bei denen w_{Si} bzw. t_{Si} systematisch variiert wurden, der in Abschnitt 4.2.2. eingeführte maximal mögliche On-Strom bei gleichzeitiger Erfüllung des I_{off} -Kriteriums bestimmt. Die dazu erforderlichen 3D Simulationen sind sehr rechenintensiv und benötigen im Schnitt etwa die 20-fache Rechenzeit vergleichbarer 2D Simulationen. Die Ergebnisse sind in Abbildung 4.6. dargestellt. Die dort gezeigten On-Ströme sind bereits korrigierte Werte der klassischen Simulationsergebnisse, wobei für jede Struktur ein Korrekturfaktor von 0.8 verwendet wurde.

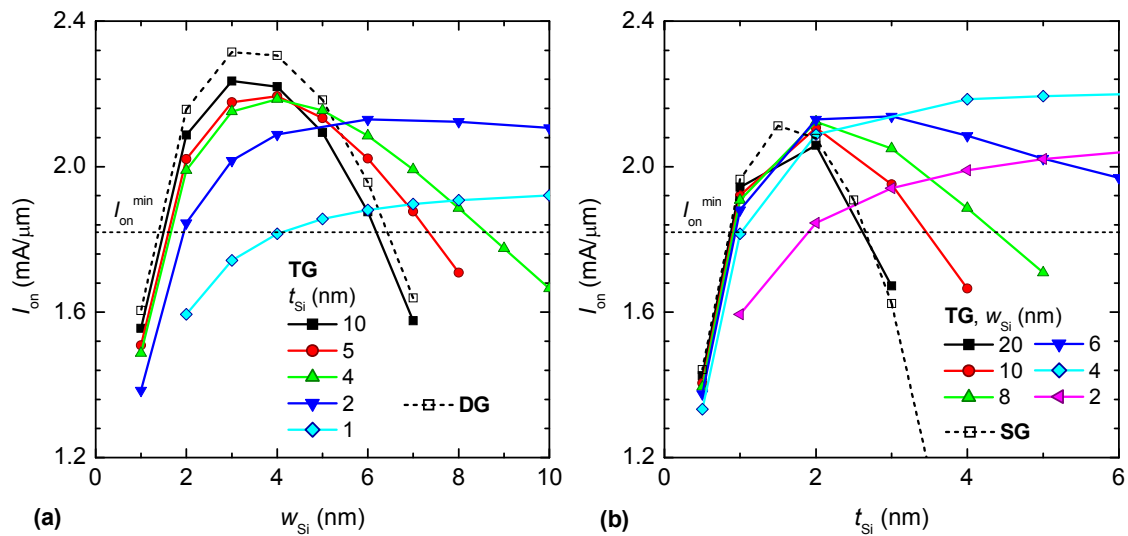


Abbildung 4.6 Maximal möglicher On-Strom von TG MOSFETs die das I_{off} -Kriterium erfüllen. (a) I_{on} als Funktion von w_{Si} mit t_{Si} als Parameter. Der Grenzfall DG bedeutet $t_{\text{Si}} \rightarrow \infty$. (b) I_{on} als Funktion von t_{Si} mit w_{Si} als Parameter. Der Grenzfall SG bedeutet $w_{\text{Si}} \rightarrow \infty$. Die On-Ströme sind auf die jeweilige effektive Gateweite normiert. Horizontale Linie: Kriterium für I_{on} .

Abbildung 4.6(a) zeigt die maximal möglichen I_{on} von TG MOSFETs mit verschiedenen Siliziumdicken als Funktion von w_{Si} . Zum Vergleich ist auch die entsprechende $I_{\text{on}}(w_{\text{Si}})$ -Kurve aus Abbildung 4.5(a) für den DG MOSFET gezeigt. Es wird deutlich, dass das Maximum der $I_{\text{on}}(w_{\text{Si}})$ -Kurven sinkt, wenn der Parameter t_{Si}

kleiner wird. Der Grund dafür sind die wachsenden Serienwiderstände R_S und R_D . Aus demselben Grund steigt auch die untere Grenze für w_{Si} . Die obere Grenze für w_{Si} steigt ebenfalls bei Verringerung von t_{Si} , da Kurzkanaleffekte zunehmend durch die Wirkung des oberen Gates unterdrückt werden. Wird t_{Si} kleiner als $t_{Si,max}$ des SG MOSFET, so ist für alle $w_{Si} > w_{Si,min}$ das I_{on} -Kriterium erfüllt. In diesem Fall ist das obere Gate für die vollständige Kontrolle des Kanalpotentials ausreichend.

Interessant ist, dass im Grenzfall DG MOSFET die obere Grenze für w_{Si} größer ist als beim TG-Design mit $t_{Si} = 10\text{nm}$. Die Ursache dafür ist das etwas schlechtere Abschaltverhalten des TG MOSFET. Hier macht sich der Einfluss des BOX bemerkbar, welches beim idealen DG MOSFET keine Rolle spielt. Beim TG MOSFET kann das Drain-Potential über das BOX auf die Rückseite der Si-Schicht wirken. Dadurch wird dort die Potentialbarriere zwischen Source und Drain verringert und die Gate-Steuerwirkung reduziert.

Dieser Effekt ist in Abbildung 4.7 deutlich sichtbar. Hier ist für einen TG MOSFET mit $w_{Si} = 6\text{nm}$ und $t_{Si} = 10\text{nm}$ im Off-Zustand die simulierte Potentialverteilung in zwei senkrechten Schnittebenen durch die Mitte der Struktur dargestellt. Auf der linken Seite ist zu sehen, wie sich das Drain-Potential innerhalb des BOX in Richtung Source ausdehnen kann. Dadurch ist rechts im Schnitt quer zum Kanal das Kanalpotential an der Stelle am größten (Barriere für Elektronen am geringsten), die am weitesten vom Gate entfernt ist, also in der Mitte der Grenzfläche zum BOX (hier: bei Tiefe $y = 10\text{nm}$ und Weite $z = 3\text{nm}$).

Obwohl dieser Effekt beim 2D DG MOSFET nicht auftritt, ist in der Praxis das DG-Konzept kein Ausweg, da t_{Si} nicht unendlich groß gemacht werden kann. Für die TG-Struktur ergibt sich aber die Forderung, dass t_{Si} möglichst groß sein sollte, wenn w_{Si} kleiner ist als $w_{Si,max}$ des DG MOSFET. Ein wirklicher Ausweg wäre nur der Übergang zur Gate-All-Around (GAA) Struktur. Diese ist aber nur mit erheblich größerem technologischen Aufwand zu realisieren.

In Abbildung 4.6(b) ist zu sehen, dass die Maxima der $I_{on}(t_{Si})$ -Kurven für das TG-Design bei Verringerung von w_{Si} zunächst größer werden und ab $w_{Si} < 4\text{nm}$ wieder sinken. Die Ursache dafür sind die beiden gegenläufigen Tendenzen bei einer Verkleinerung von w_{Si} . Einerseits verbessert sich das Abschaltverhalten (u.a. wird der BOX-Effekt zunehmend unterdrückt), andererseits werden R_S und R_D größer. Die Vergrößerung der Serienwiderstände führt gleichzeitig zur Anhebung der unteren Grenze von t_{Si} bezüglich des I_{on} -Kriteriums, das verbesserte Abschaltverhalten zur Erhöhung der oberen Grenze von t_{Si} . Wird w_{Si} kleiner als $w_{Si,max}$ des DG MOSFET, gibt

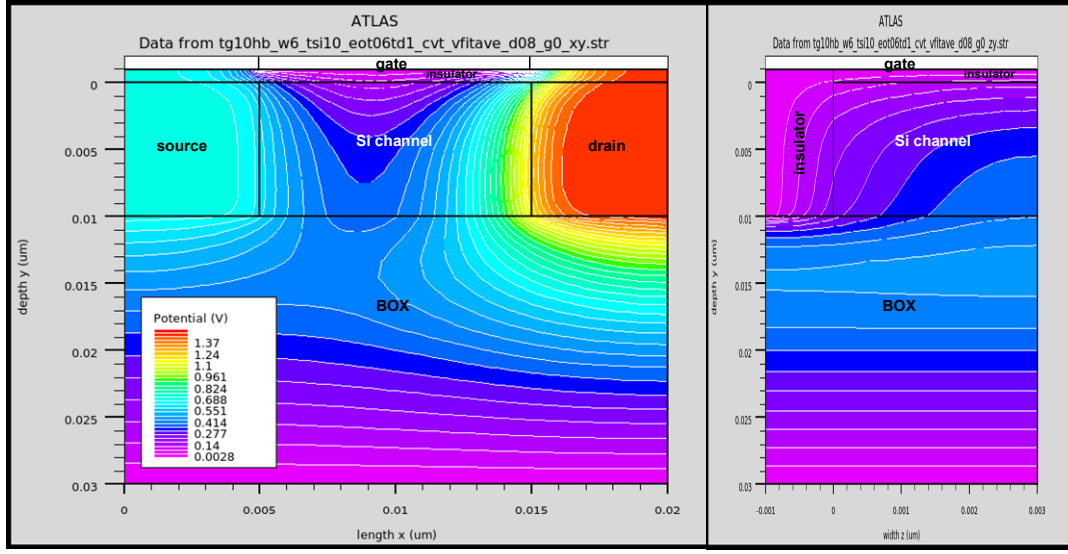


Abbildung 4.7 Simulierte Potentialverteilung in einem Tri-Gate MOSFET mit $w_{Si} = 6\text{nm}$ und $t_{Si} = 10\text{nm}$ im Off-Zustand, dargestellt in Schnittebenen durch die Mitte der Struktur. Links: Schnitt entlang des Kanals. Rechts: Schnitt quer zum Kanal.

es keine obere Grenze mehr für t_{Si} ($t_{Si,max} \rightarrow \infty$).

Vergleicht man die $I_{on}(t_{Si})$ -Kurve für das TG-Design mit $w_{Si} = 20\text{nm}$ mit der Kurve für den 2D Grenzfall SG MOSFET [Abb. 4.6(b)], so lassen sich nur geringfügige Unterschiede feststellen. Lediglich bei kleinen t_{Si} sind die I_{on} des SG MOSFET etwas größer, da hier R_S und R_D etwas kleiner sind als beim TG MOSFET.

Designspielraum für die Kanalgeometrie

Abbildung 4.8 zeigt die mit dem I_{off}/I_{on} -Kriterium ermittelte obere und untere Grenze für die TG-Querschnittsgeometrie (blaue Kurven) in der Form $w_{Si}(t_{Si})$. Die rote Kurve definiert die quantenmechanische untere Grenze für die TG-Struktur, welche nach demselben Prinzip bestimmt wurde wie $w_{Si,min}^{QM}$ bzw. $t_{Si,min}^{QM}$ für die 2D Grenzfälle DG und SG MOSFET. Die Bestimmungsgleichung hierfür lautet

$$\frac{\alpha(w_{Si,min}^{QM} + \Delta w_{Si}, t_{Si} + \Delta t_{Si})}{\alpha(w_{Si,min}^{QM}, t_{Si})} = 2, \quad (4-3)$$

wobei die Funktion α aus Gleichung (3-32) verwendet wird. Der grün schraffierte Bereich zwischen der oberen Grenze und dem quantenmechanischen Limit kennzeichnet den Designspielraum für w_{Si} und t_{Si} . TG MOSFETs mit Geometrien aus

diesem Bereich erfüllen die Kriterien für I_{on} und I_{off} , und Fluktuationen infolge von Quanteneffekten halten sich noch in akzeptablen Grenzen. TG-Designs aus dem Bereich zwischen unterer Grenze und quantenmechanischem Limit können zwar akzeptable On-Ströme liefern, aber aufgrund starker Quanteneffekte führen bereits Schwankungen von w_{Si} und t_{Si} um nur eine Atomlage zu I_{off} -Schwankungen um mehr als Faktor 2.

Zum Vergleich sind in Abbildung 4.8 auch die maximalen Schichtdicken $t_{\text{Si,max}}$ und $w_{\text{Si,max}}$ für SG bzw. DG MOSFETs markiert. TG-Designs aus dem Bereich, der von den oberen Grenzen für SG, DG und TG MOSFETs eingegrenzt wird, haben im Vergleich zu adäquaten SG- und DG-Designs größere t_{Si} bzw. w_{Si} . Das heißt aus technologischer Sicht, solche TG MOSFETs stellen geringere Anforderungen an Schichtdickenkontrolle bzw. Lithografie als vergleichbare SG bzw. DG MOSFETs.

In Tabelle 4.2 sind die wichtigsten Eigenschaften von fünf MOSFET-Varianten zusammengefasst, darunter drei ausgewählte TG MOSFET-Designs (TG1...TG3) aus dem grünen Bereich von Abbildung 4.8 sowie ein SG und ein DG MOSFET. Die effektive Gateweite der SG und DG MOSFETs ist mit unendlich angegeben, da w_{Si} bzw. t_{Si} theoretisch beliebig groß gemacht werden können. Die stufenlose Justierbarkeit von w_{eff} ist ein prinzipieller Vorteil der 2D MOSFET-Varianten. Demgegenüber sind die w_{eff} der TG MOSFETs fest eingestellt. Die einzige Möglichkeit w_{eff} von TG MOSFETs zu vergrößern, ohne das Querschnittsdesign zu ändern, besteht in der Parallelschaltung mehrerer TG-Kanäle.

Wird I_{off} über die Gateaustrittsarbeit auf $0.11 \mu\text{A}/\mu\text{m}$ eingestellt, so erreichen alle fünf MOSFET-Varianten ähnlich hohe I_{on} . Die dazu nötige Gateaustrittsarbeit Φ_{G} ist für alle Varianten ähnlich. Ein geeignetes Gatemetall wäre demnach Wolfram mit typischen Werten für die Austrittsarbeit zwischen 4.3 und 4.6eV [145]. Auch das gegenwärtig diskutierte Titanitrid TiN [146] kommt in Frage. Für TiN auf dem high-K Dielektrikum Hafniumoxid HfO_2 wurden effektive Austrittsarbeiten zwischen 4.3 und 4.5eV gemessen [147].

Die Schwellspannungen bei optimalem Φ_{G} unterscheiden sich nur unwesentlich zwischen den einzelnen MOSFET-Varianten. Die Verschiebung durch Quanteneffekte $\Delta V_{\text{th}}^{\text{QM}}$ nimmt jedoch in Richtung DG deutlich ab und ist beim SG MOSFET fast dreimal so groß wie beim DG MOSFET.

Der Kurzkanaleffekt DIBL ist bei den TG MOSFETs etwas geringer als bei SG und DG, besonders gering bei TG1. Das Abschaltverhalten ist bei allen MOSFETs mit $S < 100\text{mV}/\text{dec}$ ähnlich gut. Auch hier hat TG1 den besten Wert.

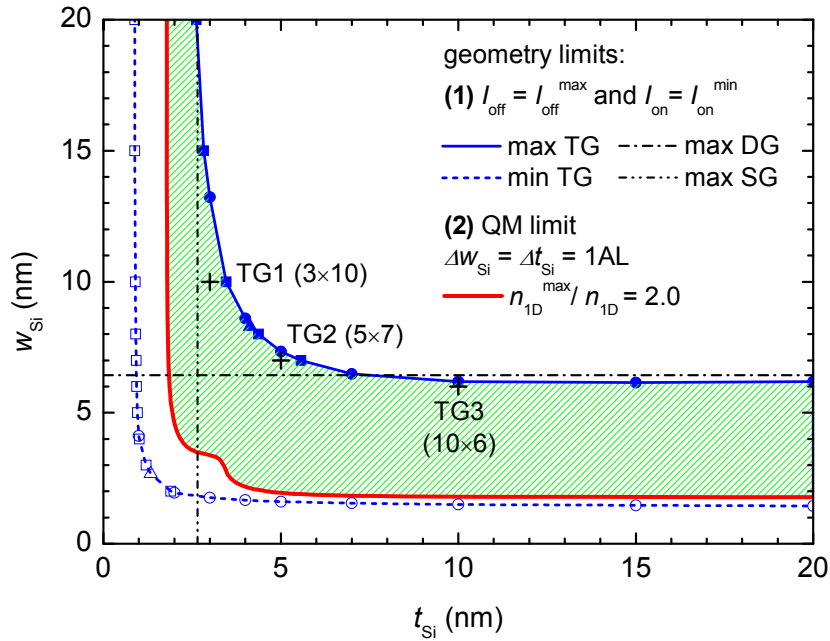


Abbildung 4.8 Obere und untere Grenzen für die Abmessungen w_{Si} und t_{Si} von TG MOSFETs. Symbole: mit dem $I_{\text{off}}/I_{\text{on}}$ -Kriterium ermittelte minimale und maximale w_{Si} und t_{Si} . Rote durchgezogene Linie ohne Symbole: quantenmechanische untere Grenze für w_{Si} und t_{Si} . Grün-schraffierter Bereich: Designspielraum für w_{Si} und t_{Si} . Strich-Punkt-Linien: $w_{\text{Si,max}}$ bzw. $t_{\text{Si,max}}$ der zweidimensionalen Grenzfälle DG und SG MOSFET.

	SG	TG1	TG2	TG3	DG
$t_{\text{Si}} \times w_{\text{Si}}$ (nm×nm)	$2.5 \times \infty$	3×10	5×7	10×6	$\infty \times 6$
w_{eff} (nm)	(∞)	16	17	26	(∞)
$I_{\text{off}}/w_{\text{eff}}$ ($\mu\text{A}/\mu\text{m}$)	0.11	0.11	0.11	0.11	0.11
$I_{\text{on}}/w_{\text{eff}}$ ($\mu\text{A}/\mu\text{m}$)	1909	1951	1876	1876	1957
Φ_{G} (eV)	4.558	4.482	4.539	4.595	4.666
V_{th} (mV)	195	183	191	193	190
$\Delta V_{\text{th}}^{\text{QM}}$ (mV)	132	127	105	71	45
DIBL (mV/V)	-176	-124	-144	-157	-182
S (mV/dec)	99.8	93.5	97.7	98.4	96.8
$R_{\text{S}}, R_{\text{D}}$ ($\Omega\mu\text{m}$)	12.8	17.1	15.5	13.9	10.7

Tabelle 4.2 Wichtige Kenngrößen ausgewählter SG, TG und DG MOSFET-Designs.

Die Source-/Drainwiderstände der untersuchten Strukturen repräsentieren die minimalen Serienwiderstände für das jeweilige Design, da sie praktisch kaum noch skalierbar sind. Kleinere R_S , R_D sind wahrscheinlich nur durch den Übergang zum Schottky-Source/Drain-Prinzip möglich, d.h. anstelle von hochdotiertem Silizium als Elektronenreservoir werden Metalle mit optimaler Austrittsarbeit verwendet. Alle R_S , R_D in Tabelle 4.2 sind kleiner als das aktuelle ITRS-Target ($60 \Omega\mu\text{m}$) [4], wobei die Widerstände der TG MOSFETs etwas größer sind als die der SG und DG MOSFETs.

Insgesamt zeigt der Vergleich, dass keine der fünf Varianten entscheidende Vorteile oder Nachteile gegenüber den anderen aufweist. Vielmehr stellt sich heraus, dass sich mit sehr verschiedenen Designs vergleichbare Eigenschaften erreichen lassen. Entscheidend für die Wahl eines bestimmten Designs ist daher die zur Verfügung stehende Technologie. Tatsächlich setzen die derzeit führenden Chiphersteller auf sehr unterschiedliche MOSFET-Konzepte für die kommenden Prozessorgenerationen [148].

4.2.4. Einfluss von Geometrieschwankungen

Ausschlaggebend für die Wahl eines bestimmten MOSFET-Designs ist nicht nur das Erreichen bestimmter elektrischer Zielgrößen, sondern auch dessen Empfindlichkeit gegenüber technologischen Prozessschwankungen. Insbesondere Toleranzen der Bauelementegeometrie können zu einer erheblichen Fluktuation der elektrischen Performance führen. Ziel der folgenden Untersuchungen ist es deshalb, den Einfluss von Schwankungen der Abmessungen t_{Si} , w_{Si} und L_G auf die elektrischen Parameter V_{th} , I_{off} und I_{on} von SG, DG und TG MOSFETs abzuschätzen. Die fünf MOSFET-Strukturen aus Tabelle 4.2 werden diesbezüglich miteinander verglichen.

Schwellschwankung

Die maximale Fluktuation der Schwellspannung aufgrund von Toleranzen der Kanalgeometrie Δt_{Si} , Δw_{Si} und ΔL_G lässt sich folgendermaßen abschätzen

$$\Delta V_{th} = \left| \frac{\partial V_{th}}{\partial t_{Si}} \Delta t_{Si} \right| + \left| \frac{\partial V_{th}}{\partial w_{Si}} \Delta w_{Si} \right| + \left| \frac{\partial V_{th}}{\partial L_G} \Delta L_G \right|. \quad (4-4)$$

Die partiellen Ableitungen in Gleichung (4-4) lassen sich entsprechend (3-34) in einen klassischen und einen quantenmechanischen Anteil zerlegen²,

$$\frac{\partial V_{th}}{\partial x} = \frac{\partial V_{th}^{CL}}{\partial x} + \frac{\partial \Delta V_{th}^{QM}}{\partial x}, \quad (4-5)$$

wobei x hier für eine beliebige geometrische Dimension steht. Unter Verwendung von (3-33) erhält man für den quantenmechanischen Term in (4-5)

$$\frac{\partial \Delta V_{th}^{QM}}{\partial x} = \frac{1}{V_T \ln 10} \left(\frac{\partial S}{\partial x} \Delta \phi^{QM} + S \frac{\partial \Delta \phi^{QM}}{\partial x} \right). \quad (4-6)$$

Für TG MOSFETs lässt sich $\Delta \phi^{QM}$ mit Gleichung (3-31) berechnen, für SG und DG MOSFETs mit Gleichung (3-35). Die Berechnung der Ableitungen von $\Delta \phi^{QM}$ erfolgt nach Gleichung (B-1) aus Anhang B. Hierbei ist zu beachten, dass $\Delta \phi^{QM}$ nicht von L_G abhängt und somit $\partial \Delta \phi^{QM} / \partial L_G = 0$ ist. Der klassische Term in (4-5) sowie S und dessen Ableitung in Gleichung (4-6) können mit Hilfe klassischer Simulationen bestimmt werden. Die Ableitungen sind näherungsweise über die Differenzenquotienten gegeben, beispielsweise ist

$$\frac{\partial V_{th}^{CL}}{\partial x} \approx \frac{V_{th}^{CL}(x + \Delta x) - V_{th}^{CL}(x - \Delta x)}{2\Delta x}. \quad (4-7)$$

Dementsprechend sind für jedes MOSFET-Design zusätzlich noch Simulationen von MOSFETs mit den Abmessungen $x + \Delta x$ und $x - \Delta x$ erforderlich. In den hier diskutierten Untersuchungen war $\Delta x = \Delta t_{Si} = \Delta w_{Si} = \Delta L_G = 1\text{nm}$, mit einer Ausnahme: beim SG MOSFET wurde ein Δt_{Si} von 0.5nm verwendet, da bei $t_{Si} = 2.5\text{nm}$ Änderungen um $\Delta t_{Si} = 1\text{nm}$ zu große Schwankungen der Kennlinien bewirken.

Abbildung 4.9 vergleicht die Auswirkung von Geometrieschwankungen auf die Schwellspannung der fünf ausgewählten MOSFET-Designs. In Abbildung 4.9(a) sind die nach Gleichung (4-5) ermittelten Ableitungen von V_{th} dargestellt, mit anderen Worten die Empfindlichkeiten von V_{th} bezüglich Fluktuationen von t_{Si} , w_{Si} und L_G . Die Empfindlichkeit gegenüber t_{Si} -Schwankungen nimmt in Richtung DG ab und ist für den DG MOSFET gleich null. Dagegen nimmt der Einfluss von w_{Si} in Richtung DG zu.

² In der weiteren Diskussion sind klassische Größen immer mit dem Index „CL“ gekennzeichnet. Variablen ohne Index bezeichnen tatsächliche Größen (unter Berücksichtigung von Quanteneffekten). Der Index „QM“ kennzeichnet Korrekturgrößen mit denen der Einfluss von Quanteneffekten berücksichtigt wird.

Beim SG MOSFET ist V_{th} unabhängig von w_{Si} . Beide Trends werden durch Quanteneffekte verstärkt. Die Empfindlichkeit gegenüber Fluktuationen von L_G ist bei allen Transistorvarianten ähnlich. Interessanterweise reduzieren Quanteneffekte die Empfindlichkeit der Schwellspannung gegenüber L_G -Schwankungen. Das lässt sich anhand von Gleichung (3-34) folgendermaßen erklären: Bei einer Verkleinerung von L_G sinkt V_{th}^{CL} aufgrund von Kurzkanaleffekten. Da aber gleichzeitig S größer wird, vergrößert sich ΔV_{th}^{QM} entsprechend Gleichung (3-33), obwohl $\Delta \phi^{QM}$ unabhängig von L_G ist. Die Vergrößerung von ΔV_{th}^{QM} dämpft die grundsätzliche Verringerung von V_{th} bei einer Gatelängenverkürzung.

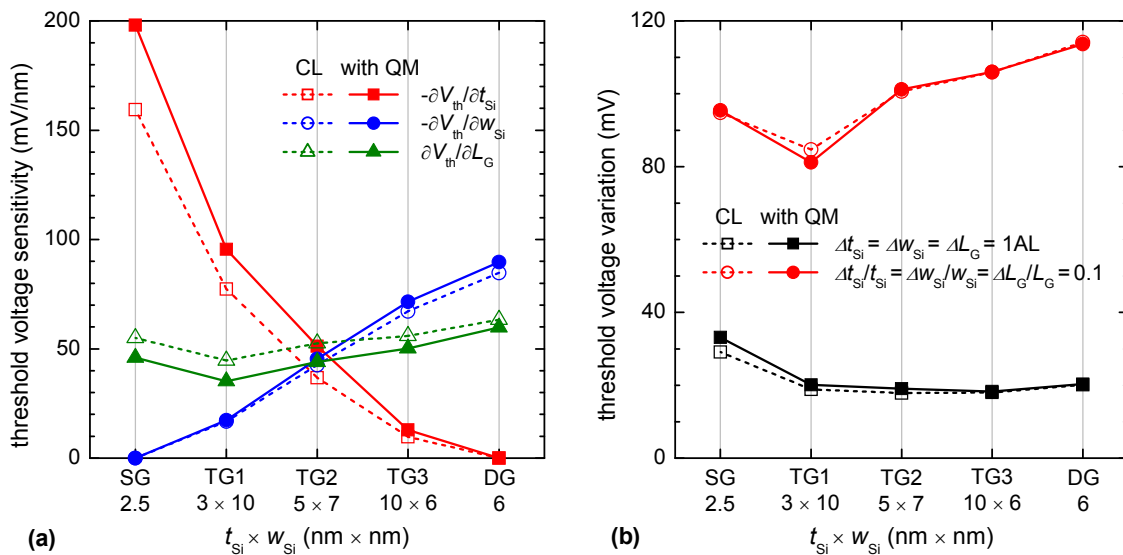


Abbildung 4.9 (a) Empfindlichkeit von V_{th} bezüglich Schwankungen von t_{Si} , w_{Si} und L_G bei ausgewählten Single-Gate, Double-Gate und Tri-Gate MOSFETs. (b) Maximale Fluktuation von V_{th} bei Schwankungen von t_{Si} , w_{Si} und L_G um eine Atomlage (schwarze Quadrate) bzw. um 10% der nominellen Abmessungen (rote Kreise). Volle Symbole: mit Quanteneffekten, offene Symbole: ohne Quanteneffekte.

Abbildung 4.9(b) zeigt für jedes MOSFET-Design die maximale Fluktuation der Schwellspannung nach Gleichung (4-4) für zwei Fälle. Im ersten Fall wird angenommen, dass alle Abmessungen um maximal eine Atomlage schwanken können, im zweiten Fall um 10% ihres Mittelwertes. In beiden Fällen haben die gegenläufigen Wirkungen der Quanteneffekte zur Folge, dass sie sich nahezu kompensieren. Im ersten Fall ist ΔV_{th} für den SG MOSFET deutlich größer als für die anderen Varianten, hält sich aber mit 33mV noch im vertretbaren Rahmen. Im zweiten Fall ist ΔV_{th} zwar für

TG1 besonders klein, für alle Varianten gilt aber $\Delta V_{th} \approx 100\text{mV}$. Das sind Schwankungen um etwa 50% des nominellen Wertes von V_{th} (vgl. Tabelle 4.2). Derartige Abweichungen sind für die meisten Anwendungen deutlich zu groß. Daraus ergibt sich die Forderung, dass Geometrietoleranzen wesentlich kleiner als 10% sein müssen. Bei den hier betrachteten Abmessungen ist das eine enorme technologische Herausforderung.

Off-Strom

Schwankungen der Schwellspannung haben exponentielle Änderungen von I_{off} zur Folge. Es ist daher sinnvoll, Schwankungen des Off-Stromes in Dekaden anzugeben. Mit Hilfe logarithmischer Ableitungen erhält man

$$\Delta \lg I_{off} = \left| \frac{\partial \lg I_{off}}{\partial t_{Si}} \Delta t_{Si} \right| + \left| \frac{\partial \lg I_{off}}{\partial w} \Delta w_{Si} \right| + \left| \frac{\partial \lg I_{off}}{\partial L_G} \Delta L_G \right|, \quad (4-8)$$

wobei $\Delta \lg I_{off}$ die maximale Fluktuation des Off-Stromes in Dekaden angibt. Damit lässt sich entsprechend

$$\frac{I_{off}}{I_{off}^{max}} = 10^{\Delta \lg I_{off}} \quad (4-9)$$

das Verhältnis von maximalem und nominellem Off-Strom bestimmen. Die Ableitungen in (4-8) lassen sich wieder in einen klassischen und einen quantenmechanischen Anteil zerlegen. Zu diesem Zweck ist es sinnvoll, von einem effektiven Subthreshold Slope

$$S = \frac{V_{th}}{\lg I_{th} - \lg I_{off}} \quad (4-10)$$

auszugehen, wobei I_{th} entsprechend (2-1) der Drainstrom ist, der bei $V_{GS} = V_{th}$ fließt. Unter Verwendung von (4-10), (3-33) und (3-34) lässt sich die Ableitung von $\lg I_{off}$ nach einer geometrischen Größe x folgendermaßen aufschreiben,

$$\frac{\partial \lg I_{off}}{\partial x} = -\frac{\partial}{\partial x} \left(\frac{V_{th}}{S} \right) = \frac{\partial \lg I_{off}^{CL}}{\partial x} - \frac{1}{V_T \ln 10} \frac{\partial \Delta \phi^{QM}}{\partial x}. \quad (4-11)$$

Der klassische Term in (4-11) lässt sich wieder mit Hilfe klassischer Simulationen bestimmen und die Ableitungen von $\Delta \phi^{QM}$ mit Gleichung (B-1).

Abbildung 4.10(a) zeigt die nach Gleichung (4-11) bestimmten Ableitungen für die fünf MOSFET-Varianten. Die Empfindlichkeiten von I_{off} bezüglich Schwankungen von t_{Si} , w_{Si} und L_{G} zeigen qualitativ ähnliche Trends wie die Empfindlichkeiten von V_{th} in Abbildung 4.9(a). Im Unterschied zu V_{th} haben die I_{off} -Empfindlichkeiten jedoch entgegen gesetzte Vorzeichen und die Empfindlichkeit von I_{off} gegenüber L_{G} wird nicht durch Quanteneffekte beeinflusst.

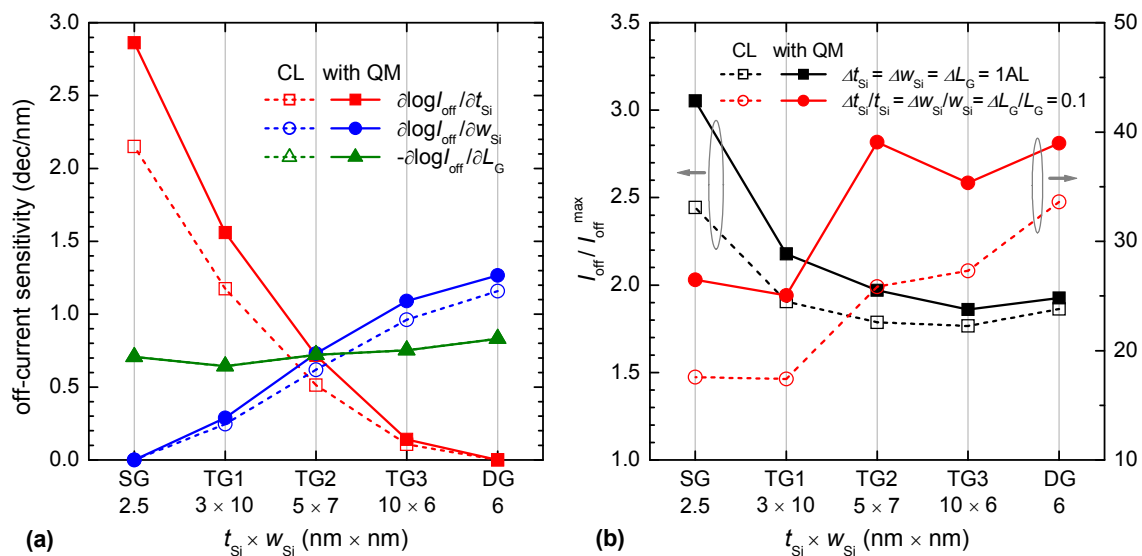


Abbildung 4.10 (a) Empfindlichkeit von I_{off} bezüglich Schwankungen von t_{Si} , w_{Si} und L_{G} . (b) Verhältnis von maximalem und nominellem Off-Strom bei Geometrie-Toleranzen von einer Atomlage bzw. von 10% der nominellen Abmessungen. Volle Symbole: mit Quanteneffekten, offene Symbole: ohne Quanteneffekte.

In Abbildung 4.10(b) ist für jedes Design das Verhältnis von maximalem und nominellem Off-Strom nach Gleichung (4-9) dargestellt. Es werden wieder die zwei Fälle von Geometrieschwankungen betrachtet, d.h. Schwankungen um eine Atomlage sowie Schwankungen um 10% der nominellen Abmessungen. Im ersten Fall nehmen I_{off} -Schwankungen in Richtung SG tendenziell zu. Generell verstärken Quanteneffekte die Fluktuation des Off-Stroms und zwar ebenfalls in Richtung SG tendenziell zunehmend. Beim SG MOSFET kann I_{off} um mehr als Faktor 3 vom nominellen Wert abweichen, aber auch bei den TG und DG MOSFETs um etwa Faktor 2 (minimal bei TG3: Faktor 1.86). Das heißt, selbst bei extrem geringen Prozesstoleranzen sind erhebliche I_{off} -Schwankungen zu erwarten. Im zweiten Fall (10%-Toleranz) vergrößern sich die Fluktuationen dann noch einmal um den Faktor 10 gegenüber dem ersten Fall.

Der Trend ist hier jedoch anders: SG und TG1 haben die geringsten I_{off} -Fluktuationen. Abweichungen vom nominellen I_{off} um mehr als Faktor 20 scheinen aber auf jeden Fall deutlich zu groß zu sein. Interessanterweise wird in der ITRS mit $\Delta L_G = 0.12 \times L_G$ gerechnet [4]. Das würde für das TG1-Design beispielsweise heißen, dass allein aufgrund der L_G -Toleranz, bereits I_{off} -Schwankungen um Faktor 6 zu erwarten sind.

On-Strom

Im Gegensatz zu Änderungen von I_{off} lassen sich Fluktuationen des On-Stroms ΔI_{on} besser im linearen Maßstab beschreiben. Die Bestimmung von ΔI_{on} erfolgt deshalb äquivalent zu ΔV_{th} nach Gleichung (4-4). Quanteneffekte beeinflussen den On-Strom über die Parameter $\Delta V_{\text{th}}^{\text{QM}}$ und c^{QM} . Während $\Delta V_{\text{th}}^{\text{QM}}$ stark von der Kanalgeometrie abhängt, kann die Geometrieabhängigkeit von c^{QM} vernachlässigt werden. Wie bereits erwähnt war in dieser Studie $c^{\text{QM}} = 0.8 = \text{const.}$ für alle Geometrien. Mit dem Ansatz

$$I_{\text{on}} = g_{\text{m,eff}}(V_{\text{DD}} - V_{\text{th}}) \quad (4-12)$$

unter Verwendung von (3-34) erhält man für die Ableitungen von I_{on} nach einer geometrischen Größe x

$$\frac{\partial I_{\text{on}}}{\partial x} = c^{\text{QM}} \frac{\partial I_{\text{on}}^{\text{CL}}}{\partial x} - g_{\text{m,eff}} \frac{\partial \Delta V_{\text{th}}^{\text{QM}}}{\partial x}. \quad (4-13)$$

Die effektive Steilheit $g_{\text{m,eff}}$ ist gegeben durch

$$g_{\text{m,eff}} = c^{\text{QM}} \frac{I_{\text{on}}^{\text{CL}}}{V_{\text{G,eff}}^{\text{CL}}} \quad (4-14)$$

und kann mit klassischen numerischen Simulationen bestimmt werden, ebenso die Ableitungen von $g_{\text{m,eff}}$ und die Ableitungen des klassischen On-Stroms $I_{\text{on}}^{\text{CL}}$. Die Größe $V_{\text{G,eff}}^{\text{CL}}$ ist die effektive Gatespannung in den klassischen Simulationen, bei der $I_{\text{on}}^{\text{CL}}$ bestimmt wird. Sie ist betragsmäßig gleich $V_{\text{DD}} - V_{\text{th}}$, jedoch unabhängig von $\Delta V_{\text{th}}^{\text{QM}}$.

In Abbildung 4.11 sind die relativen Empfindlichkeiten von I_{on} gegenüber Geometrieschwankungen sowie ΔI_{on} normiert auf I_{on} für die fünf Design-Varianten dargestellt. Die Trends der Empfindlichkeiten von I_{on} in Abbildung 4.11(a) sind qualitativ denen von V_{th} und I_{off} ähnlich. Die Vorzeichen der Empfindlichkeiten von I_{on} entsprechen denen der Empfindlichkeiten von I_{off} (vgl. Abb. 4.10a). Quanteneffekte

dämpfen die L_G -Empfindlichkeit von I_{on} , ähnlich der von V_{th} (vgl. Abb. 4.9a), da I_{on} -Fluktuationen zu einem großen Teil durch Verschiebung von V_{th} verursacht werden. Aus demselben Grund werden die t_{Si} - und w_{Si} -Empfindlichkeiten von I_{on} durch Quanteneffekte verstärkt.

In Abbildung 4.11(b) werden wie bisher die zwei Fälle von Geometrietoleranzen betrachtet. In beiden Fällen haben die gegenläufigen Wirkungen der Quanteneffekte zur Folge, dass sie sich analog zu ΔV_{th} (Abb. 4.9b) nahezu kompensieren.

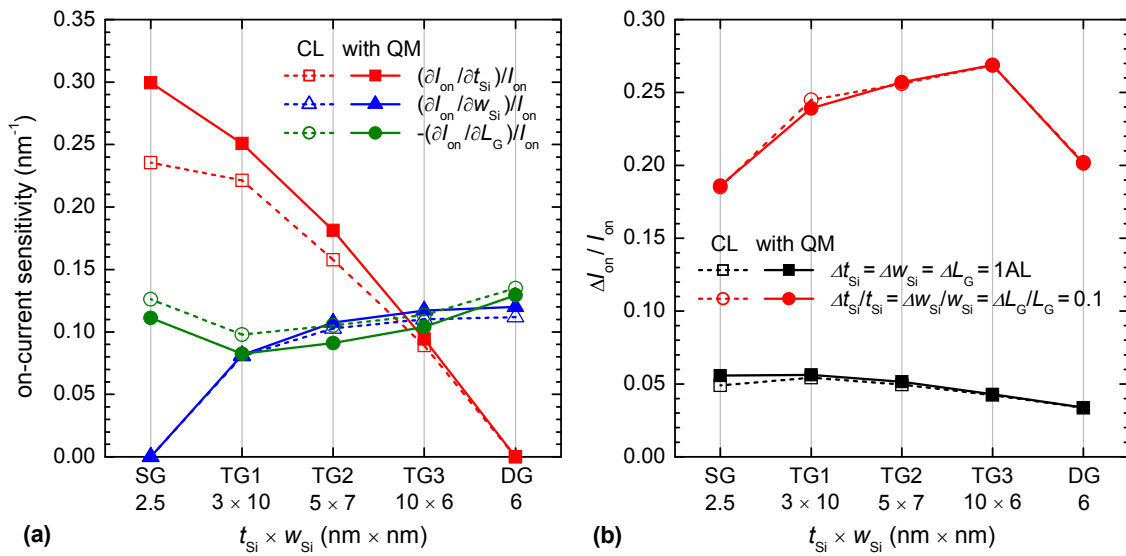


Abbildung 4.11 (a) Relative Empfindlichkeit von I_{on} bezüglich Schwankungen von t_{Si} , w_{Si} und L_G . (b) Relative Schwankung von I_{on} bei Geometrie-Toleranzen von einer Atomlage bzw. von 10% der nominellen Abmessungen. Volle Symbole: mit Quanteneffekten, offene Symbole: ohne Quanteneffekte.

Im Fall der 1AL-Toleranz werden Fluktuationen des On-Stroms in Richtung DG MOSFET kleiner. In diesem Fall sind die Abweichungen vom nominellen I_{on} aber für alle Designs relativ gering und liegen zwischen 5.6% (SG und TG1) und 3.4% (DG). Dagegen sind im Fall der 10%-Toleranz deutliche I_{on} -Schwankungen zwischen 19% (SG) und 27% (TG3) zu erwarten. In diesem Fall sind die I_{on} -Fluktuationen bei den TG MOSFETs generell stärker als bei SG oder DG MOSFETs. Solche Schwankungen von I_{on} haben entsprechend Gleichung (2-6) vor allem unterschiedliche Verzögerungszeiten der einzelnen Transistoren einer Schaltung zur Folge, was sich negativ auf die Geschwindigkeit (Taktfrequenz) der gesamten Schaltung auswirkt.

Fazit

Die Untersuchungen haben gezeigt, dass die Massenfertigung von MOSFETs mit $L_G = 10\text{nm}$ extreme Anforderungen an die Prozesskontrolle stellt, um Fluktuationen vor allem des Off-Stroms in vertretbaren Grenzen zu halten. Die Toleranzen der Kanalgeometrie sollten im Schnitt deutlich unter 10% liegen, d.h. eher im Bereich der Dicke einer Atomlage.

Der Vergleich von Single-Gate, Double-Gate und Trigate MOSFETs liefert kein bevorzugtes MOSFET-Design für diesen Gatelängenbereich. Diese Aussage weicht von den Ergebnissen anderer Studien ab, die das DG-Konzept [146], [149]-[150] gegenüber dem TG-Konzept favorisieren. Dagegen ist die Empfehlung der vorliegenden Studie, das Design so zu wählen, dass mit der zur Verfügung stehenden Technologie die Performance-Schwankungen möglichst gering sind. Dementsprechend hat auch das SG SOI-Konzept durchaus seine Berechtigung. Das Fazit lautet also: *Für jede Technologie gibt es ein optimales MOSFET-Design.* Als eine Methode zur Abschätzung der zu erwartenden Performance-Schwankungen kann die hier vorgestellte Kombination aus klassischer numerischer Simulation und analytischen Quantenkorrekturen hilfreich sein.

4.3. Einfluss von parasitären Widerständen

Die bisherigen Untersuchungen konzentrierten sich auf das Design und die elektrischen Eigenschaften des inneren Transistors, wobei nur die minimalen intrinsischen Bahnwiderstände der Source- und Draingebiete berücksichtigt wurden³. Die Performance realer MOSFETs wird jedoch auch wesentlich durch parasitäre Widerstände bestimmt. Zu R_S und R_D tragen beispielsweise die Kontaktwiderstände der Metall/Halbleiterübergänge sowie die Widerstände der Metallleitbahnen bei. Außerdem sind die Source-/Drainkontakte normalerweise nicht senkrecht bis zum BOX durchmetallisiert, wie bisher angenommen, sondern liegen flach auf dem Silizium. Bei einer solchen Anordnung ist der Widerstand der Bahngebiete zwischen ohmschen Kontakt und Kanal deutlich größer als in der bisher betrachteten idealisierten Struktur.

Die folgenden Untersuchungen beschäftigen sich mit der Wirkung größerer parasitärer Serienwiderstände auf die Gleichstromeigenschaften von MOSFETs mit 10nm Gatelänge, insbesondere auf den On-Strom. Es geht dabei um die grundsätzlichen

³ vgl. Diskussion zu R_S , R_D in Tab. 4.2 am Ende von Abschnitt 4.2.3.

Auswirkungen von R_S und R_D , d.h. alle infrage kommenden Widerstandskomponenten werden zusammengefasst als kompakte ohmsche Widerstände an den Source- und Drainkontakten betrachtet.

4.3.1. Theoretische Betrachtungen

Wirkung von R_S und R_D

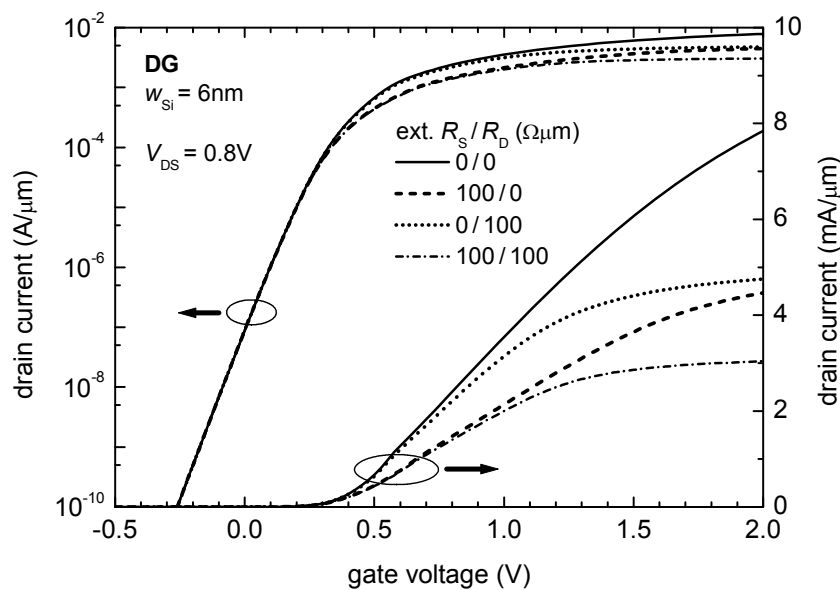


Abbildung 4.12 Transferkennlinien (ohne Quantenkorrektur) eines DG MOSFETs mit und ohne externe R_S und/oder R_D im logarithmischen und linearen Maßstab.

Abbildung 4.12 zeigt simulierte Transferkennlinien eines DG MOSFET mit $w_{Si} = 6\text{nm}$ mit und ohne äußere Widerstände an Source und/oder Drain von jeweils $100\Omega\mu\text{m}$. Während man im Subthresholdbereich (logarithmischer Maßstab) keinen Einfluss von R_S oder R_D erkennen kann, ist deren Wirkung im eingeschalteten Zustand (linearer Maßstab) offensichtlich. Der Sourcewiderstand R_S bewirkt im Wesentlichen eine gleichmäßige Verringerung der Steilheit. Demgegenüber hat R_D bis $V_{GS} \approx 1\text{V}$ nur eine geringfügige Verkleinerung der Steilheit zur Folge, bewirkt danach jedoch eine Sättigung von I_D . Wirken beide Widerstände gleichzeitig, lassen sich die von R_S bzw. R_D dominierten Kennlinienbereiche deutlich unterscheiden.

Die Wirkung der Serienwiderstände lässt sich anhand des Ersatzschaltbildes in Abbildung 4.13 gut verstehen. Es besteht aus dem inneren Transistor und den externen

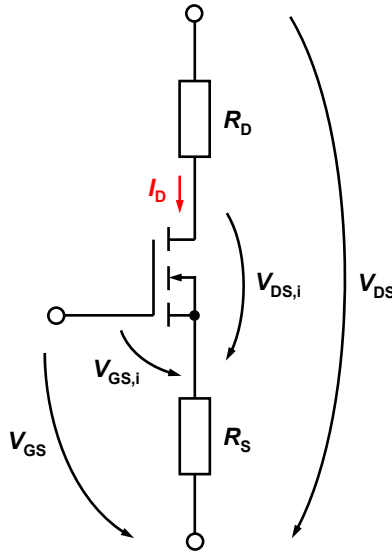


Abbildung 4.13 Gleichstromersatzschaltbild eines MOSFET mit parasitären Widerständen.

(parasitären) Widerständen R_S und R_D . Bei von außen angelegten Spannungen V_{GS} und V_{DS} werden am inneren Transistor die Spannungen

$$V_{GS,i} = V_{GS} - I_D R_S, \quad (4-15)$$

und

$$V_{DS,i} = V_{DS} - I_D (R_S + R_D) \quad (4-16)$$

wirksam.

Gleichung (4-16) liefert eine Erklärung für das Sättigungsverhalten der Transferkennlinie. Da bei $V_{DS,i} = 0$ kein Strom mehr fließen kann, gilt grundsätzlich $I_D < V_{DS}/(R_S + R_D)$. Das heißt, selbst wenn man V_{GS} unendlich groß machen würde, könnte I_D diese Grenze nicht überschreiten.

Anhand von Gleichung (4-15) lässt sich verstehen, warum durch R_S die Steilheit reduziert wird. Wegen des Spannungsabfalls über R_S muss für eine bestimmte Stromänderung die äußere Gate-Source-Spannung V_{GS} stärker geändert werden als $V_{GS,i}$, d.h. stärker als die Gate-Source-Spannung ohne R_S . Mit dem Ansatz

$$g_m = \frac{dI_D}{dV_{GS,i}} \cdot \frac{dV_{GS,i}}{dV_{GS}} = g_{m,i} \frac{dV_{GS,i}}{dV_{GS}} \quad (4-17)$$

unter Verwendung von (4-15) erhält man für die Steilheit den Ausdruck

$$g_m = \frac{g_{m,i}}{1 + g_{m,i}R_S}, \quad (4-18)$$

wobei $g_{m,i}$ die Steilheit des inneren Transistors ist. Gleichung (4-18) sagt aus, dass die außen messbare Steilheit eines Transistors umso geringer ist, je größer dessen parasitärer Sourcewiderstand ist. Nur im Grenzfall $g_{m,i}R_S \ll 1$ spielt der Sourcewiderstand keine Rolle.

Quantenkorrektur bei großen Serienwiderständen

Wie wirken sich Quanteneffekte auf den On-Strom von MOSFETs mit großen Serienwiderständen aus? Beim bisher betrachteten inneren Transistor konnte die Verringerung der Steilheit bzw. des Drainstroms des inneren Transistors mit dem Korrekturfaktor $c^{QM} = n_S^{QM}/n_S^{CL}$ (vgl. Abschnitt 3.4.4) berücksichtigt werden. Über den Ansatz (4-12) lässt sich auch der Einfluss von R_S auf I_D berücksichtigen, wenn man für $g_{m,eff}$ einen Ausdruck der Form (4-18) verwendet,

$$I_D = \frac{g_{m,eff,i}}{1 + g_{m,eff,i}R_S} (V_{GS} - V_{th}), \quad (4-19)$$

wobei

$$g_{m,eff,i} = I_D / (V_{GS,i} - V_{th}) \quad (4-20)$$

die effektive Steilheit des inneren Transistors ist. Quanteneffekte können über den Korrekturfaktor c^{QM} entsprechend

$$g_{m,eff,i} = c^{QM} g_{m,eff,i}^{CL} \quad (4-21)$$

berücksichtigt werden, wobei $g_{m,eff,i}^{CL}$ mit Hilfe klassischer Simulationen bestimmt werden kann.

Im Allgemeinen ist der Korrekturfaktor c^{QM} für einen betrachteten MOSFET nicht konstant, sondern hängt von dessen Arbeitspunkt ab, genauer gesagt von $V_{GS,i} - V_{th}$. Für die hier relevanten Strukturen (DG mit $w_{Si} > 2\text{nm}$) ist jedoch laut Abbildung 3.35 das Verhältnis n_S^{QM}/n_S^{CL} im interessanten Bereich der effektiven Gatespannung zwischen 0.3V und 0.6V nahezu konstant. Deshalb kann der für 0.6V bestimmte Wert aus Abbildung 3.36 für die jeweilige Struktur verwendet werden.

Aus (4-19) und (4-21) folgt, dass mit zunehmenden R_S der relative Einfluss von Quanteneffekten auf I_D geringer wird. Das wird deutlich, wenn man das Verhältnis der Ströme mit und ohne Quantenkorrektur

$$\frac{I_D}{I_D^{CL}} = c^{QM} \frac{1 + g_{m,eff,i}^{CL} R_S}{1 + c^{QM} g_{m,eff,i}^{CL} R_S} \quad (4-22)$$

betrachtet. Für sehr große R_S geht dieses Verhältnis gegen 1.

Für die praktische Anwendung der Quantenkorrektur entsprechend (4-19) – (4-21) ist es sinnvoll, die etwas unhandliche Größe $g_{m,eff,i}$ zu eliminieren. Durch Einsetzen von (4-15), (4-20) und (4-21) in Gleichung (4-19) erhält man

$$I_D = c^{QM} I_D^{CL} \frac{V_{GS,eff}}{V_{GS,eff} + (c^{QM} - 1) R_S I_D^{CL}}, \quad (4-23)$$

wobei $V_{GS,eff} = V_{GS} - V_{th}$ ist. Im Grenzfall kleiner R_S folgt aus Gleichung (4-23) die in Abschnitt 4.2 verwendete Quantenkorrektur $I_D \approx c^{QM} I_D^{CL}$.

4.3.2. Auswirkung von Serienwiderständen auf I_{on}

Nun bleibt noch die Frage zu klären, wie stark die On-Ströme extrem skaliert MOSFETs durch Serienwiderstände reduziert werden.

Zu diesem Zweck wurden Simulationen des DG MOSFET mit $w_{Si} = 6\text{nm}$ und des SG MOSFET mit $t_{Si} = 2.5\text{nm}$ durchgeführt, wobei äußere Widerstände an den Source- und Drainkontakten variiert wurden. Dabei war stets $R_S = R_D$, was die Realität gut widerspiegelt, da MOSFETs für digitale Anwendungen normalerweise symmetrisch aufgebaut sind.

Abbildung 4.14 zeigt die maximal möglichen I_{on} bei gleichzeitiger Erfüllung des I_{off} -Kriteriums mit und ohne Quantenkorrektur nach Gleichung (4-23) als Funktion von R_S und R_D . In dieser Darstellung wird der starke Einfluss parasitärer Widerstände auf die On-Ströme der betrachteten MOSFET-Strukturen deutlich. Es genügt bereits eine Verdopplung der internen (minimalen) Bahnwiderstände (vgl. Tabelle 4.2) um das I_{on} -Kriterium zu verfehlen. Mit Serienwiderständen, die gerade die ITRS-Vorgaben erfüllen, d.h. $R_S = R_D = 60\Omega\mu\text{m}$, erreichen beide Transistorstrukturen nur etwa 80% des I_{on} -Targets. Dieses Ergebnis unterstreicht den enormen Einfluss parasitärer Widerstände auf die Performance stark skaliert MOSFETs. Die Minimierung externer Widerstände

ist die Voraussetzung dafür, überhaupt einen nennenswerten Performancegewinn durch die weitere Skalierung des Si MOSFET erzielen zu können [151]. Die Ergebnisse aus Abbildung 4.14 legen den Schluss nahe, dass die Vorgaben der ITRS bezüglich I_{on} und I_{off} für High-Performance MOSFETs mit $L_G = 10\text{nm}$ aufgrund externer Source- und Drainwiderstände möglicherweise nicht gleichzeitig erreichbar sind.

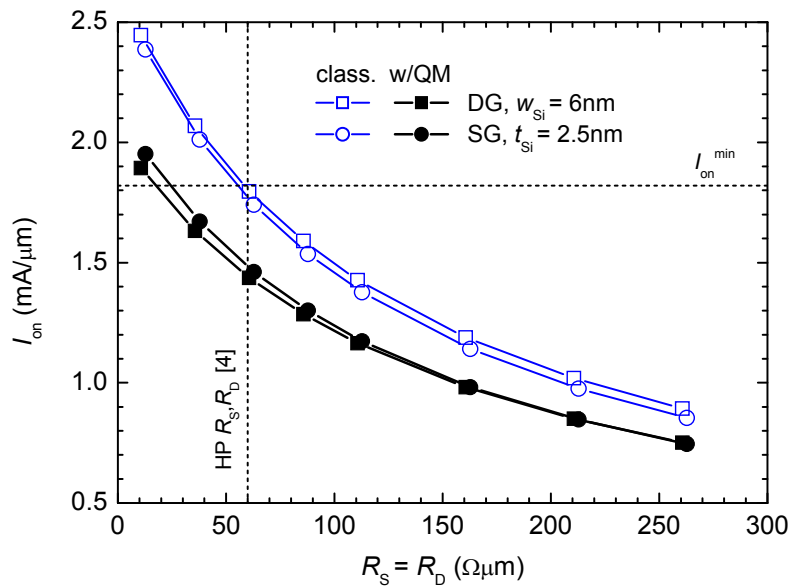


Abbildung 4.14 Maximal möglicher On-Strom von DG und SG MOSFETs bei gleichzeitiger Erfüllung des I_{off} -Kriteriums als Funktion von R_S und R_D . Offene Symbole: klassische Simulationsergebnisse; volle Symbole: Ergebnisse mit Quantenkorrektur nach Gleichung (4-23). Gestrichelten Linien markieren Zielgrößen für I_{on} (horizontale Linie) sowie R_S und R_D (vertikale Linie). Letztere ist das aktuelle ITRS-Target für High-Performance Logik MOSFETs mit $L_G = 10\text{nm}$ [4].

Kapitel 5

Nanometer-MOSFETs für Hochfrequenzanwendungen

5.1. Einführung

Obwohl Silizium im Vergleich zu anderen Halbleitermaterialien, wie z.B. den III-V Verbindungshalbleitern (GaAs, InP), nicht gerade optimale Eigenschaften für Hochfrequenzanwendungen besitzt, hat sich der Silizium MOSFET seit Anfang der 2000er Jahre als ein wichtiges Bauelement für Anwendungen im unteren Gigahertzbereich etabliert [35]. Zwar haben III-V FETs im Vergleich zu Si MOSFETs bei gleicher Gatelänge normalerweise deutlich höhere Grenzfrequenzen, die kontinuierliche Skalierung des Si MOSFET hat jedoch zu einem enormen Anstieg seiner intrinsischen Geschwindigkeit sowie zu einer deutlichen Verkleinerung parasitärer Elemente geführt. Daher kommt der Si MOSFET nun für viele Anwendungen infrage, die bis vor wenigen Jahren noch ausschließlich Transistoren aus III-V Verbindungshalbleitern vorbehalten waren.

Die entscheidenden Vorteile des Si MOSFET als Hochfrequenzbauelement gegenüber anderen Transistorarten sind seine kostengünstige und hochentwickelte Technologie, sowie seine inhärente Kompatibilität mit CMOS-Logik. Die Integrierbarkeit von Hochfrequenz-MOSFETs (HF MOSFETs) und CMOS-Logik auf einen Chip trägt wesentlich zur Kostenreduzierung bei. Das macht den Si MOSFET besonders attraktiv für Massenmarktprodukte im Bereich Mobilkommunikation, wo der Kostenfaktor eine entscheidende Rolle spielt. Tatsächlich ist RF (radio frequency) CMOS heute die dominierende Technologie für mobile Internetverbindungen mit kurzer Reichweite, wie WLAN (wireless local area network) oder Bluetooth [4]. Mit zunehmender Verbesserung der Hochfrequenzeigenschaften wird der Si MOSFET auch für Anwendungen im höheren Gigahertzbereich (20...100GHz) interessant [4].

In Abbildung 5.1 ist die von der ITRS prognostizierte Entwicklung der Grenzfrequenzen f_T und f_{max} von Performance-RF/Analog MOSFETs dargestellt. Die

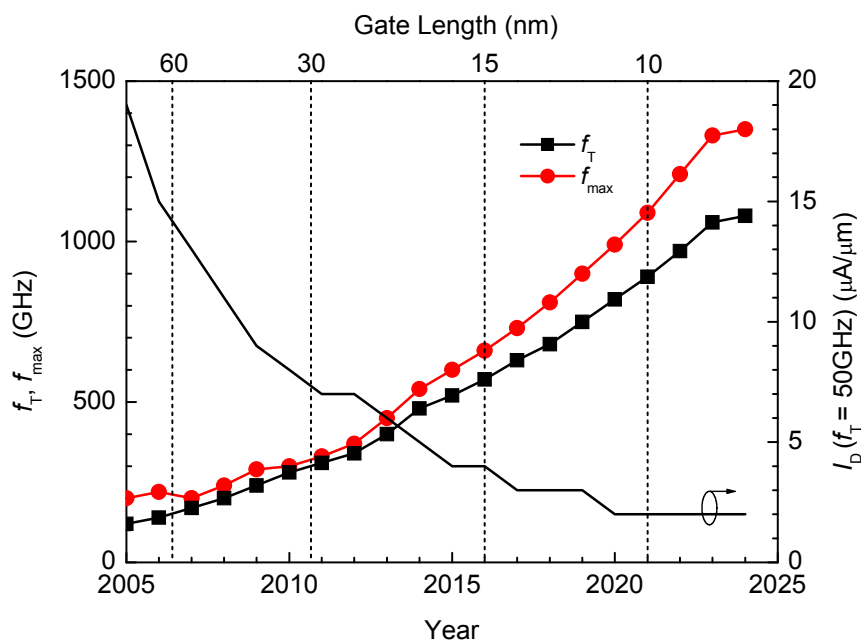


Abbildung 5.1 ITRS-Targets für Performance-RF/analog MOSFETs bis zum Jahr 2024 [4]. Gezeigt ist die prognostizierte Entwicklung von f_T und f_{\max} sowie des erforderlichen Drainstroms für ein f_T von 50GHz. Die Jahreszahl (untere x-Achse) gibt die erwartete Einführung in die Produktion an. Vertikale Linien kennzeichnen die Gatelänge (obere x-Achse) der RF MOSFETs, die zum entsprechenden Zeitpunkt in die Produktion eingeführt werden.

Grenzfrequenzen aktueller MOSFETs mit Gatelängen um 30nm liegen bei etwa 300GHz [152]-[153]. Für das Jahr 2021 wird mit der Einführung von RF MOSFETs mit 10nm Gatelänge in die Massenproduktion gerechnet, deren f_{\max} die Grenze von einem Terrahertz übersteigen soll. Derart hohe Grenzfrequenzen sind für viele Anwendungen zwar nicht zwingend erforderlich – die dargestellten ITRS-Targets gelten für Anwendungen im Frequenzbereich von 0.4 bis 10GHz – haben aber entscheidende qualitative Verbesserungen zur Folge. So ist beispielsweise das Signal-Rausch-Verhältnis SNR (signal-to-noise ratio) am Ausgang eines Low-Noise-Amplifier (LNA) umso größer (das Eigenrauschen des LNA umso geringer), je größer das f_T der verwendeten Transistoren ist [34], [154]. In Radioempfängern bestimmt das Eigenrauschen von LNAs die minimale Signalstärke, welche zuverlässig demoduliert werden kann. Diese ist ein Schlüsselfaktor für Reichweite und Leistungsaufnahme von mobilen Kommunikationssystemen [34].

Falls die extrem hohen Werte für f_T und f_{\max} in Abbildung 5.1 tatsächlich erreicht werden können, kommen Si MOSFETs selbstverständlich auch für neue Anwendungen bei wesentlich höheren Arbeitsfrequenzen infrage.

Neben den Grenzfrequenzen ist in Abbildung 5.1 eine weitere wichtige Zielgröße dargestellt: der Drainstrom, welcher für ein f_T von 50GHz erforderlich ist. Dieser sinkt mit fortschreitender Skalierung. Das bedeutet, die Transistoren können für das gleiche h_{21} bzw. U mit einem geringeren Drainstrom betrieben werden. Das kann dazu genutzt werden, die DC-Verlustleistung von Verstärkerschaltungen zu verringern. Auf diese Art lässt sich beispielsweise die Standby-Verlustleistung von Mobiltelefonen reduzieren.

In den folgenden Abschnitten werden Untersuchungen zu Hochfrequenzeigenschaften von MOSFETs mit einer Gatelänge von 10nm vorgestellt. Diese Untersuchungen konzentrieren sich auf die zweidimensionalen MOSFET-Designs aus Abbildung 4.2, d.h. auf SG und DG MOSFETs. Auf die Untersuchung von Tri-Gate MOSFETs wurde wegen des enormen Rechenaufwandes von 3D Simulationen mit Kleinsignalanalyse verzichtet.

Zwei Fragen bilden den Hindergrund der nachfolgenden Diskussionen:

- (1) Gelten für optimale Hochfrequenzeigenschaften ähnliche Designregeln wie für optimale Gleichstromeigenschaften? Oder anders gefragt: *Muss ein guter HF MOSFET auch ein guter Digital-MOSFET sein?*
- (2) Unter welchen Voraussetzungen sind die Vorgaben der ITRS für f_T und f_{\max} erreichbar? Insbesondere soll geklärt werden, wie groß parasitäre Widerstände und Kapazitäten maximal sein dürfen.

In Abschnitt 5.2 wird zunächst wird die Abhängigkeit der Grenzfrequenzen vom Design des inneren Transistors betrachtet. Anschließend wird in Abschnitt 5.3 die Wirkung parasitärer Widerstände und Kapazitäten diskutiert.

5.2. Design des inneren Transistors

In den folgenden Abschnitten wird die Auswirkung grundlegender Designparameter wie Kanalquerschnitt und äquivalente Oxiddicke auf die Hochfrequenzeigenschaften von SG und DG MOSFETs mit 10nm Gatelänge (vgl. Abb. 4.2) untersucht. Externe Strukturen, wie die Source-/Drainanschlussgebiete oder das Gatedesign werden dabei nicht betrachtet. Solche Strukturen haben die Wirkungen parasitärer Widerstände und Kapazitäten, welche in Abschnitt 5.3 diskutiert werden.

Die ITRS-Targets für Transistorgeometrie und Betriebsspannung von Performance-RF MOSFETs orientieren sich an den Vorgaben für Low-Standby-Power (LSTP) Logik [4].

Im Vergleich zu High-Performance (HP) Logik sind bei LSTP die äquivalente Oxiddicke und die Betriebsspannung etwas größer. In den hier vorgestellten Simulationen wurde eine Betriebsspannung von 1V angenommen. Die äquivalente Oxiddicke betrug standardmäßig 1nm.

In Hinblick auf Frage (1) wurden nicht nur die Grenzfrequenzen der verschiedenen MOSFET Designs betrachtet sondern auch die DC-Parameter I_{on} und I_{off} . Mit ATLAS wurden die Transferkennlinien von SG und DG MOSFETs mit $L_G = 10\text{nm}$ bei $V_{DS} = 1\text{V}$ simuliert. In jedem Gleichstromarbeitspunkt wurde eine Kleinsignalanalyse (S^3A) für eine Frequenz von 40GHz durchgeführt sowie f_T und f_{max} mit den Gleichungen (2-15) und (2-16) bestimmt. Aus den Transferkennlinien wurde außerdem, wie in Abschnitt 4.2.2 beschrieben, der maximal mögliche On-Strom bei gleichzeitiger Erfüllung des I_{off} -Kriteriums extrahiert. Im Unterschied zu Kapitel 4 beziehen sich nun jedoch alle Aussagen auf LSTP-Logik. Bei LSTP MOSFETs dieser Größe darf der Off-Strom maximal $10\text{pA}/\mu\text{m}$ betragen [4]. Das sind 4 Größenordnungen weniger als beispielsweise bei HP-Logik erlaubt sind. Die Zielgröße für I_{on} richtet sich nach der gewählten Betriebsspannung. Bei $V_{DD} = 1\text{V}$ ist ein On-Strom um $1.3\text{mA}/\mu\text{m}$ angemessen. Die Zielgrößen für f_T und f_{max} betragen 870GHz bzw. 1160GHz.

5.2.1. Kanalquerschnitt

Betrachten wir zunächst die Gleichstromeigenschaften. In Abbildung 5.2 sind die maximal möglichen I_{on} von SG und DG MOSFETs bei gleichzeitiger Erfüllung des I_{off} -Kriteriums als Funktion der Siliziumdicke dargestellt. Gezeigt sind Werte mit und ohne Quantenkorrektur. Die Zielgröße für den On-Strom, I_{onLSTP} , sowie die mit Gleichung (4-2) definierte, durch Quanteneffekte begrenzte minimale Schichtdicke $w_{Si,min}^{QM}$ sind in Abbildung 5.2 besonders markiert. Es wird deutlich, dass sowohl SG als auch DG MOSFETs On-Ströme liefern können, die größer sind als I_{onLSTP} . Um das zu erreichen, muss jedoch beim SG MOSFET, im Gegensatz zum DG MOSFET, die Siliziumdicke kleiner sein als $w_{Si,min}^{QM}$. Demnach ist das DG-Konzept für LSTP-Anwendungen dem SG-Konzept eindeutig vorzuziehen. Damit aber ein DG MOSFET das I_{on} -Kriterium erfüllen kann, darf w_{Si} nicht größer als 4nm sein. Das ist deutlich geringer als beispielsweise bei HP Logik möglich wäre (vgl. Abb. 4.5a).

Kommen wir nun zu den Hochfrequenzeigenschaften dieser Transistoren. In Abbildung 5.3 sind die simulierten f_T von DG MOSFETs mit verschiedenen w_{Si} als Funktion der

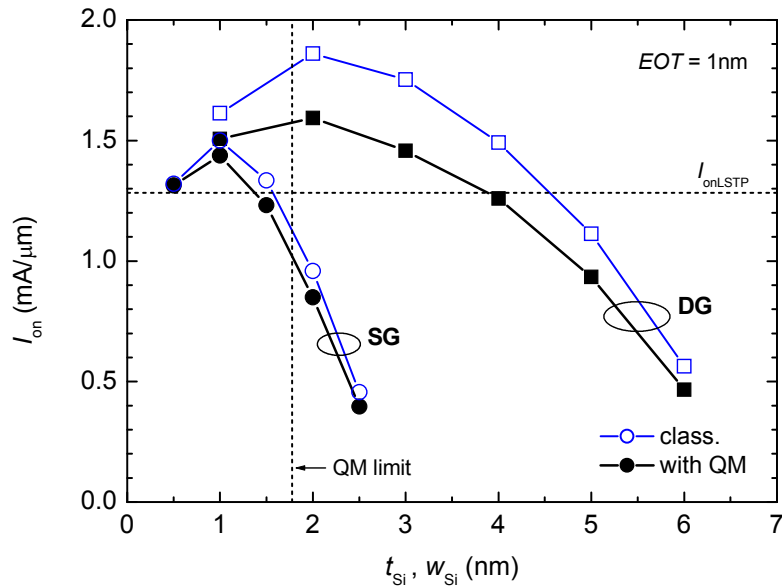


Abbildung 5.2 Maximal möglicher On-Strom bei gleichzeitiger Erfüllung des I_{off} -Kriteriums als Funktion der Siliziumdicke von SG und DG MOSFETs mit $L_G = 10$ nm für LSTP Anwendungen. Gezeigt sind jeweils die I_{on} mit und ohne Quantenkorrektur. Für die Quantenkorrektur wurde der variable Korrekturfaktor n_s^{QM}/n_s^{CL} aus Abb. 3.36 verwendet. Gestrichelte Linien – horizontal: Zielgröße für I_{on} , vertikal: durch Quanteneffekte definierte minimale Schichtdicke.

Gatespannung dargestellt. Das f_T -Maximum liegt für alle Transistoren bei etwa 2.5THz. Das ist beinahe dreimal so hoch wie die Zielgröße von 870GHz. Der Si MOSFET ist also zumindest prinzipiell in der Lage die ITRS-Targets zu erfüllen. Diese extrem hohen f_T -Werte kennzeichnen jedoch lediglich die intrinsische Geschwindigkeit der Transistoren. Ob sie auch tatsächlich erreicht werden können, hängt in erster Linie von der Größe (unvermeidbarer) parasitärer Elemente ab.

Bei genauer Betrachtung fällt auf, dass die Maximalwerte für f_T mit zunehmender Siliziumdicke leicht ansteigen. Der Grund dafür sind die mit Vergrößerung von w_{Si} kleiner werdenden (intrinsischen) Source- und Drainwiderstände, wodurch entsprechend Gleichung (2-20) f_T größer wird.

Außerdem vergrößert sich mit zunehmendem w_{Si} der Bereich von V_{GS} mit sehr großen f_T -Werten. Die Ursache dafür ist ein Zusammenspiel verschiedener Effekte. In Richtung kleiner V_{GS} ist der entscheidende Grund die Verringerung der Schwellspannung mit zunehmendem w_{Si} , d.h. die Verschiebung der Transferkennlinien auf der Spannungsachse (vgl. Abb. 4.3).

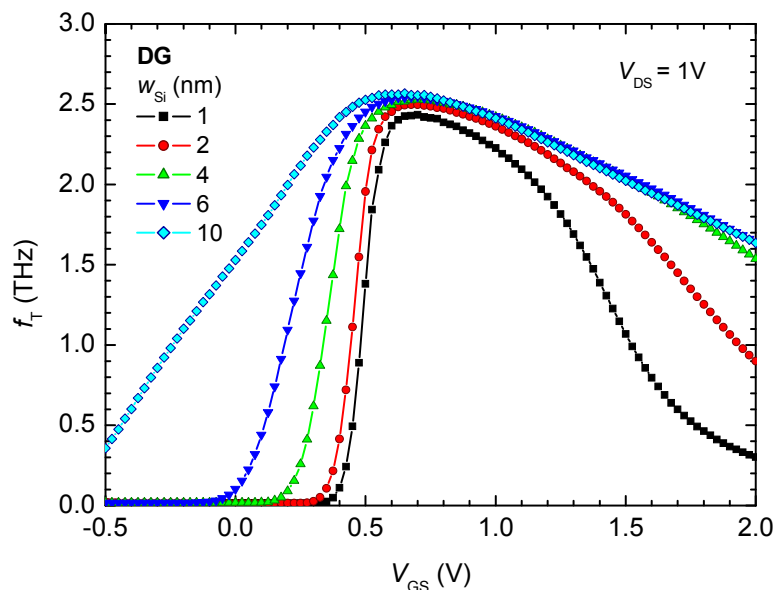


Abbildung 5.3 Simulierte Transitfrequenz von DG MOSFETs mit $L_G = 10nm$ und unterschiedlicher Siliziumdicke als Funktion der Gatespannung.

In Richtung großer V_{GS} ist die Wirkung der Source- und Drainbahnwiderstände entscheidend. Diese sorgen entsprechend (4-15) und (4-16) für eine Verschiebung des Gleichstromarbeitspunktes am inneren Transistor¹. Das führt zu einer Verringerung der Steilheit des inneren Transistors und somit, entsprechend Gleichung (2-20), zu einer Verringerung von f_T . Bei kleinen w_{Si} ist dieser Effekt wegen der größeren Serienwiderstände stärker.

Zusätzlich verstärkt wird das Absinken von f_T in Richtung großer V_{GS} durch zwei sekundäre Effekte, die Vergrößerung der Bahnwiderstände sowie die Zunahme des Drainleitwerts g_{DS} mit ansteigendem Drainstrom. Die Bahnwiderstände werden größer, da mit zunehmendem Strom aufgrund des sich vergrößernden elektrischen Feldes in Transportrichtung die Elektronenbeweglichkeit in den Bahngebieten abnimmt [Gleichung (3-6)]. Bei kleinen w_{Si} ist dieser Effekt stärker ausgeprägt, da dort die Änderung des elektrischen Feldes größer ist als bei großen w_{Si} . Bei $w_{Si} = 2nm$ beispielsweise vergrößert sich R_S von $V_{GS} = 1V$ zu $V_{GS} = 2V$ um den Faktor 1.5. Zum Vergleich: bei $w_{Si} = 6nm$ vergrößert sich R_S nur um Faktor 1.07.

Verantwortlich für das Ansteigen des Drainleitwerts ist die Verschiebung des Gleichstrom-arbeitspunktes am inneren Transistor. Beim DG MOSFET mit $w_{Si} = 2nm$

¹ Gemeint ist hier der Transistor ohne Bahnwiderstände.

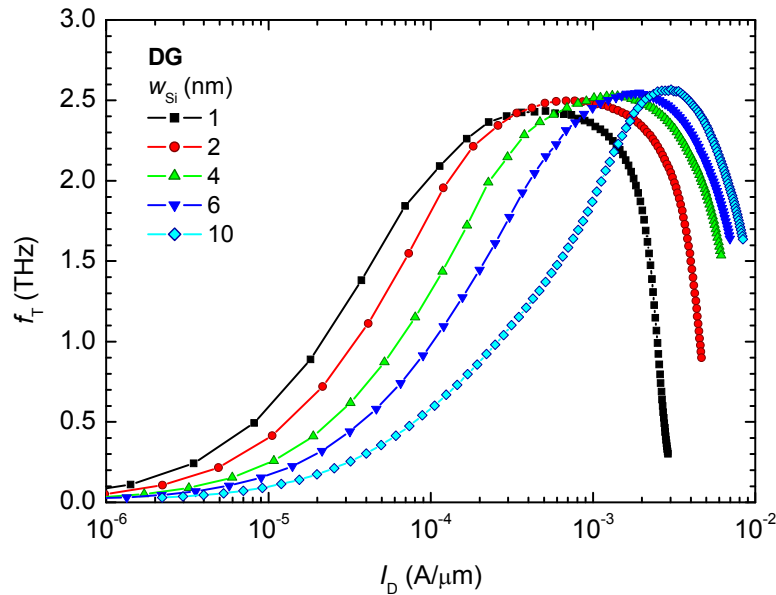


Abbildung 5.4 Transitfrequenz von DG MOSFETs mit $L_G = 10$ nm und unterschiedlicher Siliziumdicke als Funktion des Drainstroms.

vergrößert sich g_{DS} zwischen $V_{GS} = 1$ V und $V_{GS} = 2$ V um mehr als Faktor 10, bei $w_{Si} = 6$ nm nur um Faktor 2. Entsprechend Gleichung (2-20) ergibt das zusammen mit den größeren R_S und R_D ein noch stärkeres Absinken von f_T mit zunehmender Gatespannung bei kleinen w_{Si} .

Man könnte nun behaupten, ein schlechter Digital-MOSFET ist für Hochfrequenzanwendungen besonders gut geeignet. Betrachtet man jedoch f_T als Funktion des Drainstroms (Abb. 5.4), ergibt sich ein etwas anderes Bild. Es zeigt sich, dass das f_T -Maximum bei umso kleineren I_D erreicht wird, je kleiner w_{Si} ist. Das bedeutet, je besser die Gleichstrom-eigenschaften eines Transistors sind, desto geringer ist die DC-Verlustleistung bei gleichem f_T .

In Abbildung 5.5 sind die maximalen f_T und f_{max} von SG und DG MOSFETs als Funktion der Siliziumdicke dargestellt. Um SG und DG MOSFETs mit gleich großen Bahnwiderständen direkt vergleichen zu können, wurden die Grenzfrequenzen des DG MOSFET gegen $w_{Si}/2$ aufgetragen². Es ist zu sehen, dass SG und DG MOSFETs ähnlich hohe f_T und f_{max} zeigen, wobei die Grenzfrequenzen des SG MOSFET bei gleichen Bahnwiderständen durchweg etwas geringer sind. Letzteres ist auf das größere senkrechte Feld im Kanal von SG MOSFETs im Vergleich zu DG MOSFETs

² Nach Gleichung (4-1) sind die Bahnwiderstände eines DG MOSFET doppelt so groß wie die eines SG MOSFET mit gleicher Siliziumdicke.

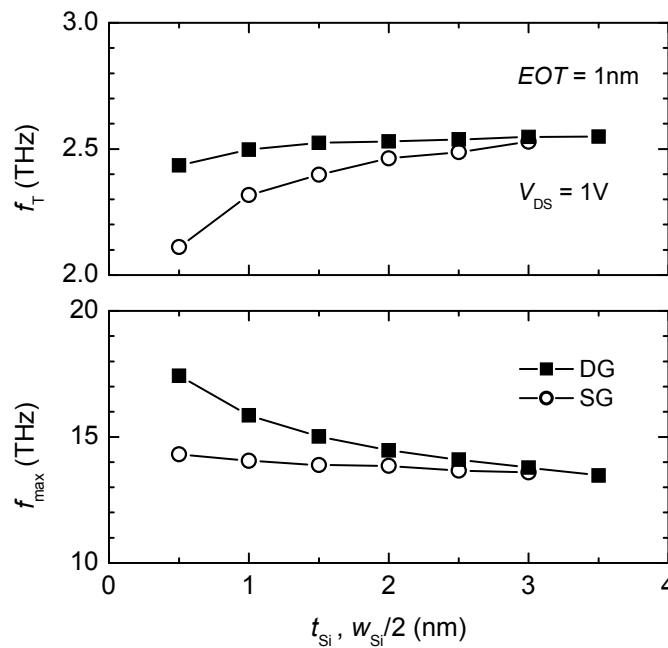


Abbildung 5.5 Maximale f_T und f_{max} als Funktion von t_{Si} (SG MOSFET) bzw. $w_{Si}/2$ (DG MOSFET).

zurückzuführen. Aufgrund seines symmetrischen Aufbaus ist im DG MOSFET das senkrechte Feld effektiv geringer. Dadurch ist die effektive Elektronenbeweglichkeit im Kanal des SG MOSFET kleiner als im Kanal des DG MOSFET, insbesondere bei kleinen Siliziumdicken. Die Simulationen ergeben im DG MOSFET mit $w_{Si} = 1\text{nm}$ eine um 13% höhere effektive Beweglichkeit als im SG MOSFET mit $t_{Si} = 0.5\text{nm}$.

Für beide MOSFET-Typen liefern die Simulationen extrem hohe f_{max} -Werte um 15THz. Das ist das Sechsfache der simulierten f_T . Der Grund dafür ist der ideale Gatekontakt der Simulationsstrukturen mit $R_G = 0$ [vgl. Gleichung (2-21)]. Im Gegensatz zu f_T sinkt f_{max} mit zunehmender Schichtdicke. Ursache dafür ist die Vergrößerung von g_{DS} aufgrund zunehmender Kurzkanaleffekte. Beim DG MOSFET, zum Beispiel, verfünffacht sich g_{DS} zwischen $w_{Si} = 2\text{nm}$ und $w_{Si} = 6\text{nm}$. Beim SG MOSFET sinkt f_{max} weniger stark mit zunehmender Schichtdicke aufgrund des stärkeren Anstiegs von f_T [Gleichung (2-21)].

In Abbildung 5.6 ist für SG und DG MOSFETs der Drainstrom bei $f_T = 50\text{GHz}$, $I_D(50\text{GHz})$, als Funktion der Siliziumdicke dargestellt. Bei allen Strukturen sind das nur wenige $\mu\text{A}/\mu\text{m}$, das heißt der jeweilige Gleichstromarbeitspunkt befindet sich im Subthresholdbereich. Das verspricht für viele Gigahertzanwendungen eine extrem geringe DC-Verlustleistung.

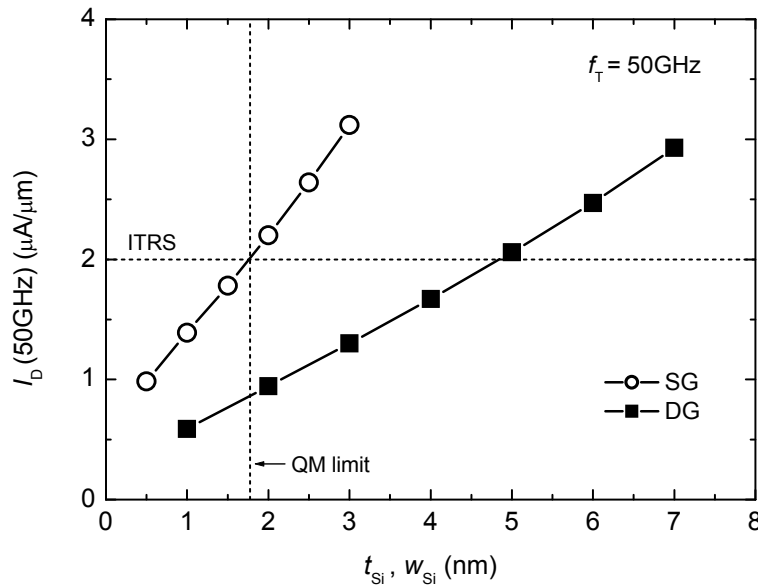


Abbildung 5.6 Erforderlicher Drainstrom für ein f_T von 50GHz als Funktion der Siliziumdicke von SG und DG MOSFETs. Gestrichelte Linien – horizontal: ITRS-Target [4], vertikal: durch Quanteneffekte definierte minimale Schichtdicke.

Sowohl bei SG als auch bei DG MOSFETs steigt $I_D(50GHz)$ nahezu linear mit der Schichtdicke an. Dieser Anstieg ist beim SG MOSFET etwa doppelt so stark wie beim DG MOSFET. Das ITRS-Target von $2\mu A/\mu m$ wird mit dem SG-Konzept bei $t_{Si} \approx 1.8nm$ und mit DG-Konzept bei $w_{Si} \approx 5nm$ überschritten. Das bedeutet, zur Erfüllung dieses Targets darf die Siliziumdicke des DG MOSFET um den Faktor 2.8 größer sein als die des SG MOSFET. Die Siliziumdicke des SG MOSFET darf wiederum höchstens so groß sein wie die durch Gleichung (4-2) definierte minimale Schichtdicke. Bei solch dünnen Schichten ist mit erheblichen I_{off} -Schwankungen zu rechnen. Das DG-Konzept hat hier also deutliche Vorteile gegenüber dem SG-Konzept.

Zum besseren Verständnis dieser Ergebnisse sind einige einfache analytische Betrachtungen hilfreich. Im Subthresholdbereich ist der Einfluss der Serienwiderstände marginal, da g_m und g_{DS} sehr klein sind. Es ist daher sinnvoll, von Gleichung (2-22) bzw. Gleichung (3-13) auszugehen. Betrachten wir zunächst g_m . Im Subthresholdbereich hängt der Drainstrom exponentiell von V_{GS} ab. Entsprechend Abbildung 5.7 gilt

$$\lg \frac{I_D}{I_1} = \frac{V_{GS} - V_1}{S}, \quad (5-1)$$

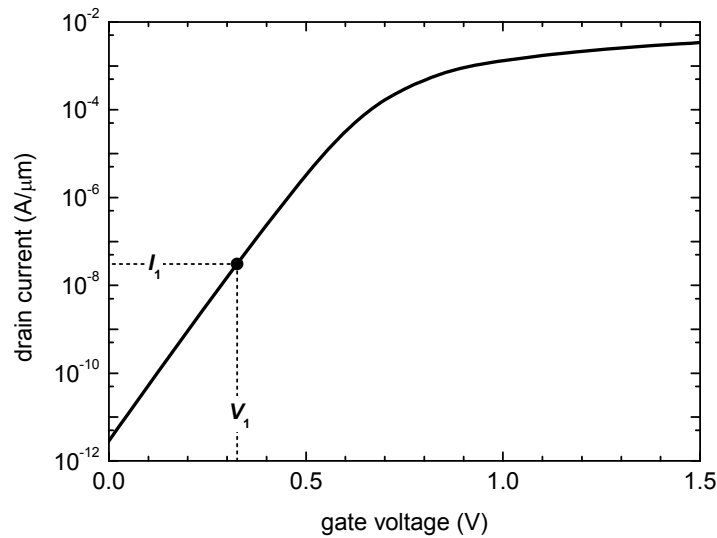


Abbildung 5.7 Transferringekennlinie in logarithmischer Darstellung. In dieser Darstellung ist die Kennlinie im Subthresholdbereich eine Gerade mit dem Anstieg $1/S$. Der Strom I_1 ist ein beliebiger Referenzstrom und V_1 die dazugehörige Gatespannung.

wobei I_1 und V_1 ein frei wählbarer Referenzstrom und die dazugehörige Gatespannung sind, die beliebig nahe bei den betrachteten I_D bzw. V_{GS} liegen können.

Löst man Gleichung (5-1) nach I_D auf, ergibt die Ableitung nach V_{GS}

$$\frac{dI_D}{dV_{GS}} = g_m = \frac{I_D}{S} \ln 10. \quad (5-2)$$

Im Subthresholdbereich ist g_m also proportional zum Drainstrom und wird durch Kurzkanaleffekte ($S > 60\text{mV/dec}$) reduziert. Setzt man (5-2) in Gleichung (3-13) ein, erhält man f_T als Funktion von I_D , oder eben auch I_D als Funktion von f_T für den Subthresholdbereich.

$$I_D = \frac{2\pi}{\ln 10} \cdot S \cdot C_{GG} \cdot f_T, \quad (5-3)$$

d.h. im Subthresholdbereich sind Drainstrom und f_T direkt proportional zueinander. Ein interessanter Aspekt von Gleichung (5-3) ist, dass sowohl S als auch C_{GG} nahezu unabhängig von V_{GS} bzw. I_D sind. Beide Größen hängen in erster Linie nur von der Transistorgeometrie ab. Sind Kurzkanaleffekte relevant, kommt bei S der Einfluss von V_{DS} noch hinzu.

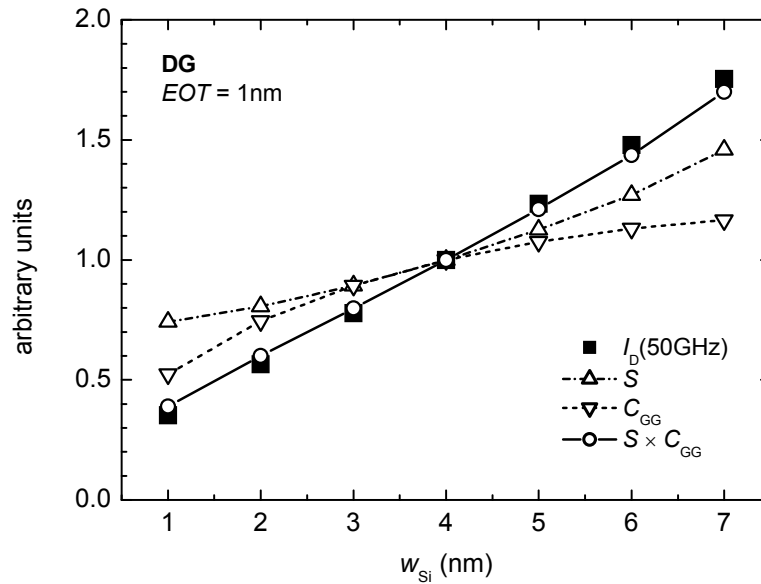


Abbildung 5.8 Drainstrom $I_D(50GHz)$, Subthreshold Slope S , Gatekapazität C_{GG} sowie das Produkt $S \times C_{GG}$ als Funktion der Siliziumdicke.

In Abbildung 5.8 sind die für den DG MOSFET bei $f_T = 50GHz$ simulierten I_D , S und C_{GG} sowie das Produkt $S \times C_{GG}$ als Funktion von w_{Si} dargestellt. Für den Vergleich wurden alle Größen auf ihren jeweiligen Wert bei $w_{Si} = 4nm$ normiert. Die Abhängigkeit des Drainstroms und die des Produkts $S \times C_{GG}$ von der Schichtdicke sind nahezu identisch, was die Gültigkeit von Gleichung (5-3) bestätigt.

Laut Abbildung 5.8 werden sowohl S als auch C_{GG} mit wachsendem w_{Si} größer. Die Zunahme von S ist auf die stärker werdenden Kurzkanaleffekte zurückzuführen (vgl. Abschnitt 4.2.2.). Die Vergrößerung von C_{GG} ist auf den ersten Blick weniger einleuchtend. Grundsätzlich kann man sich die Gatekapazität C_{GG} im Subthresholdbereich als eine Kombination aus Oxidkapazität C_{ox} und der Kapazität des vollständig verarmten Kanalgebietes C_d vorstellen. Der Einfluss beweglicher Ladungsträger im Kanal ist vernachlässigbar gering. Man könnte nun annehmen, dass C_d wie beim Langkanal-Bulk-MOSFET von der Dicke der verarmten Schicht w_d entsprechend [29] $C_d = \epsilon_{Si}/w_d$ (flächenbezogen) abhängt. Auf die DG-Struktur übertragen hieße das, $C_d = 2\epsilon_{Si}/w_{Si}$ ist die Verarmungskapazität pro Gatefläche³. Dann müsste C_{GG} aber mit zunehmender Schichtdicke kleiner werden. Für den extrem

³ Bacarani et al. bezeichnen den Ausdruck $C_d = 4\epsilon_{Si}/w_{Si}$ als die Verarmungskapazität der DG-Struktur pro Flächeneinheit [112].

skalierten DG MOSFET ist diese Vorstellung von C_d offensichtlich falsch. Hier muss die zweidimensionale Elektrostatik der Transistorstruktur betrachtet werden. Die Situation lässt sich adäquat modellieren, wenn man das Kanalgebiet durch ein ideales Dielektrikum mit der DEK von Silizium ersetzt und die hochdotierten Source- und Draingebiete als ideale Metallelektroden betrachtet. In Abbildung 5.9 ist eine derartige Modellstruktur für den DG MOSFET zu sehen. Die Kapazität einer solchen Anordnung hängt nicht von der angelegten Spannung ab, sondern nur von der Geometrie. Aus Gründen der Symmetrie genügt deshalb die Betrachtung eines Viertels der Struktur. Das sind im Prinzip zwei im rechten Winkel zueinander angeordnete Kondensatorplatten. Es ist klar, dass die Kapazität dieser Anordnung von der Fläche der Elektroden abhängt. Je größer die Elektrodenfläche, d.h. je größer w_{Si} oder L_G ist, desto größer muss auch die Kapazität sein.

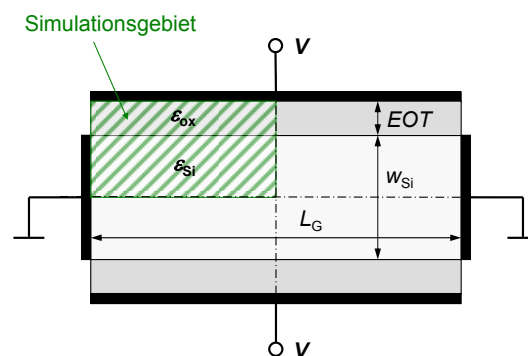


Abbildung 5.9 Modellstruktur zur Beschreibung der Gatekapazität eines DG MOSFET im Subthresholdbereich. Die vollständig verarmte Si-Schicht wird durch ein ideales Dielektrikum ersetzt. Die Source-/Draingebiete werden durch Metallelektroden beschrieben. Aus Symmetriegründen genügt die Betrachtung eines Viertels der Struktur.

Das Sättigungsverhalten der $C_{GG}(w_{Si})$ -Kurve in Abbildung 5.8 lässt sich mit dieser Modellvorstellung ebenfalls gut verstehen. Je länger eine der Platten gemacht wird, umso geringer wirkt sich jede weitere Verlängerung auf die Gesamtkapazität aus, da die hinzukommende Plattenfläche immer weiter von der gegenüberliegenden Kondensatorplatte entfernt ist.

In Abbildung 5.10 ist die mit ATLAS simulierte Kapazität der Modellstruktur aus Abbildung 5.9 im Vergleich zum C_{GG} des DG MOSFET als Funktion von w_{Si} zu sehen. Ebenfalls gezeigt sind analytisch berechnete Werte, wobei C_{GG} (normiert auf die effektive Gateweite) als Reihenschaltung von $C_{ox} = \epsilon_{ox}L_G/EOT$ und $C_d = 2\epsilon_{Si}L_G/w_{Si}$

betrachtet wurde. Für alle w_{Si} stimmt die Kapazität der Modellstruktur sehr gut mit C_{GG} des DG MOSFET überein, während das analytische Modell besonders für kleine w_{Si} vollkommen andere Werte liefert. Die im Vergleich zum DG MOSFET etwas größeren Kapazitäten der Modellstruktur sind auf die vereinfachte Modellierung der Source- und Draingebiete mit Metallelektroden zurückzuführen. Im DG MOSFET werden die Ränder der hochdotierten Gebiete ebenfalls geringfügig verarmt, so dass im Modell der effektive Abstand der Elektroden tatsächlich etwas größer sein müsste und damit die Kapazität etwas geringer. Auf jeden Fall bietet die Modellstruktur einen Ansatzpunkt, um zu einem korrekten analytischen Modell für C_{GG} und somit auch für f_{T} im Subthresholdbereich zu gelangen.

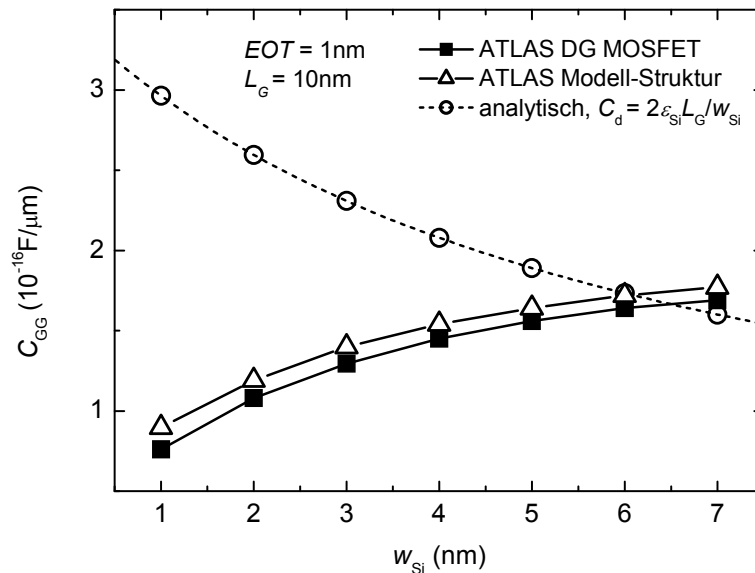
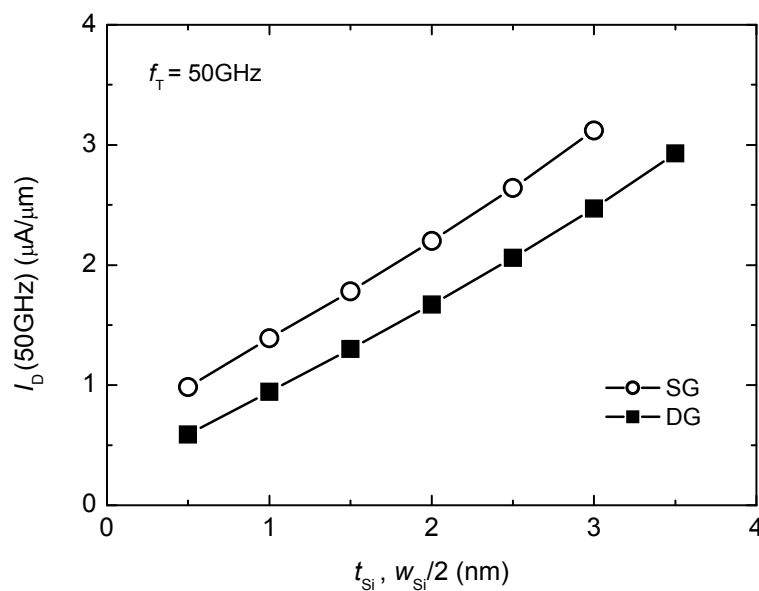


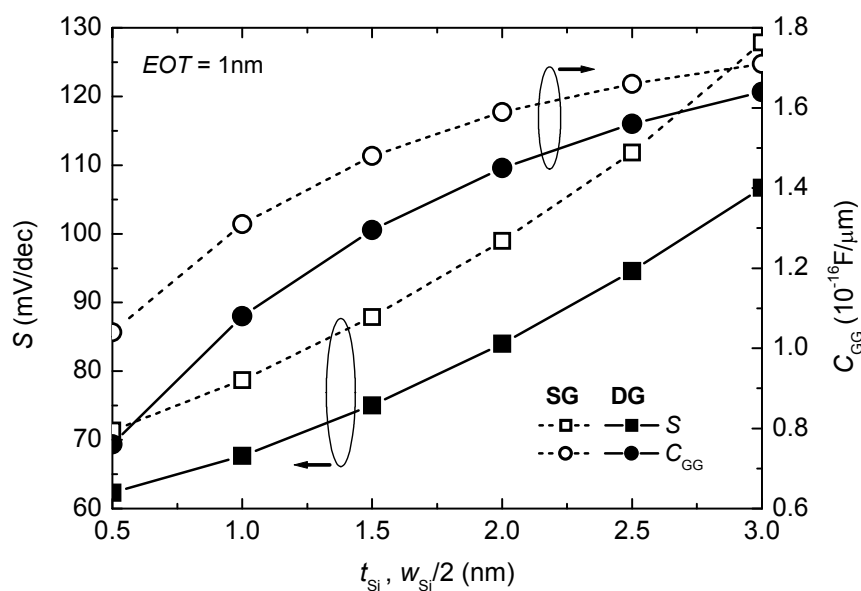
Abbildung 5.10 Gatekapazität des DG MOSFET im Subthresholdbereich (gefüllte Quadrate) im Vergleich zur Kapazität der Modellstruktur aus Abb. 5.9 (offene Dreiecke) als Funktion von w_{Si} . Ebenfalls gezeigt ist die mit einem analytischen Modell berechnete Gatekapazität (offene Kreise).

Kommen wir noch einmal auf den Vergleich von SG und DG MOSFETs hinsichtlich $I_{\text{D}}(50\text{GHz})$ in Abbildung 5.6 zurück. Aufgrund der symmetrischen Struktur des DG MOSFET muss jedes Gate nur das Potential in der dazugehörigen Hälfte der Kanalschicht steuern. Im SG MOSFET wiederum ist das eine Gate allein für die Steuerung der gesamten Kanalschicht zuständig. Aus diesem Grund ist das etwa doppelt so starke Ansteigen von $I_{\text{D}}(50\text{GHz})$ mit der Schichtdicke beim SG MOSFET nicht verwunderlich. Für einen Vergleich der Drainströme von SG und DG MOSFET ist es daher sinnvoll, die Werte bei $t_{\text{Si}} = w_{\text{Si}}/2$ zu vergleichen. In Abbildung 5.11(a) ist zu

sehen, dass bei vergleichbaren Schichtdicken der $I_D(50\text{GHz})$ des SG MOSFET deutlich größer ist als der des DG MOSFET.



(a)



(b)

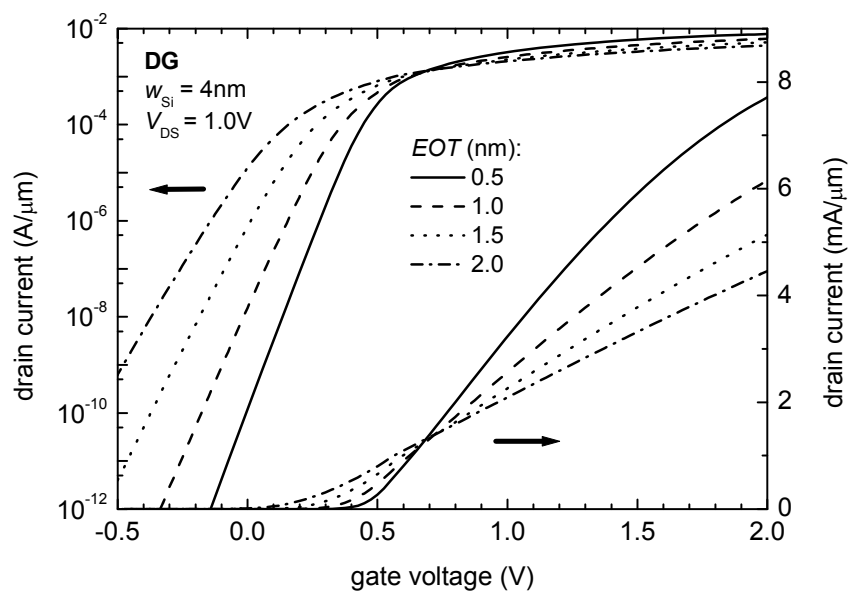
Abbildung 5.11 (a) Drainstrom $I_D(50\text{GHz})$ sowie (b) Subthreshold Slope S und Gatekapazität C_{GG} als Funktion von t_{Si} (SG MOSFET) bzw. $w_{\text{Si}}/2$ (DG MOSFET).

Betrachtet man S und C_{GG} in Abbildung 5.11(b), so sind auch diese Werte beim SG MOSFET durchweg höher. Während S beim SG MOSFET etwas stärker mit t_{Si} ansteigt als S beim DG MOSFET mit $w_{Si}/2$, nähert sich C_{GG} des SG MOSFET zu größeren Schichtdicken hin den Werten des DG MOSFET an. Beides hängt mit dem Einfluss des vergrabenen Oxids beim SG MOSFET zusammen. Der in Abschnitt 4.2.3. diskutierte BOX-Effekt ist die Ursache für das schlechtere Abschaltverhalten (größeres S) beim SG MOSFET. Dessen Einfluss wird mit Vergrößerung von t_{Si} stärker. Darüber hinaus stellt das BOX zusammen mit dem Rückseitensubstrat eine zusätzliche Kapazität dar, die eine Vergrößerung von C_{GG} bewirkt. Betrachtet man das Rückseitensubstrat als ideale Metallelektrode auf Sourcepotential, so hat man im Vergleich zur halben Modellstruktur des DG MOSFET aus Abbildung 5.9 bei der SG-Struktur eine größere effektive Elektrodenfläche. Es ist klar, dass der Einfluss dieser zusätzlichen Elektrodenfläche umso geringer ist je dicker die Siliziumschicht ist, da zum einen der prozentuale Anteil an der Gesamtfläche abnimmt und zum anderen der Abstand von der Gateelektrode zunimmt.

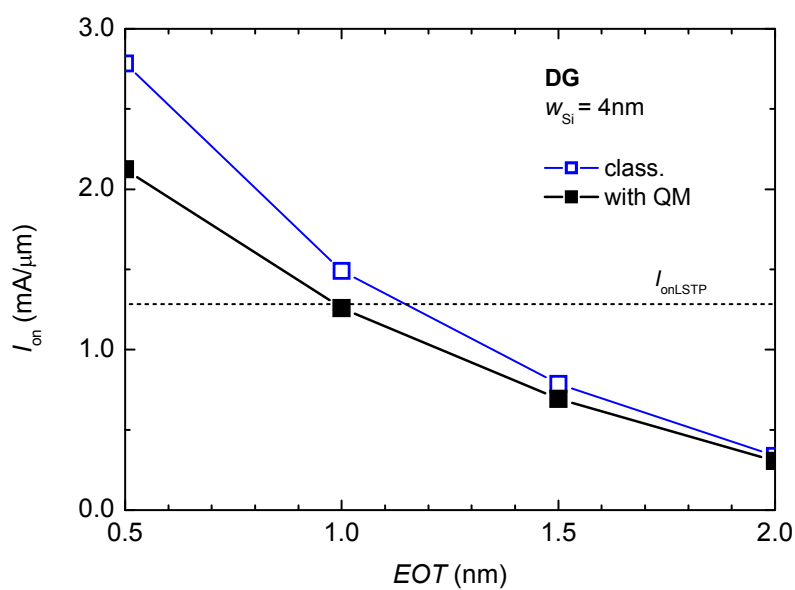
5.2.2. Äquivalente Oxiddicke

Abbildung 5.12 zeigt den Einfluss der äquivalenten Oxiddicke auf die Gleichstromeigenschaften von DG MOSFETs. Die Transferkennlinien in Abbildung 5.12(a) machen deutlich, dass sich mit der Verkleinerung von EOT die Wirkung von Kurzkanaleffekten verringert. Die Schwellspannung nimmt zu und S wird kleiner. Im On-Zustand führt die Vergrößerung der Gatekapazität zu einer deutlichen Erhöhung der Steilheit. Sowohl das verbesserte Abschaltverhalten als auch die größere Steilheit haben eine Vergrößerung des maximal möglichen On-Stroms [Abb 5.12(b)] zur Folge.

Verglichen mit dem Einfluss auf I_{on} wirkt sich eine Änderung von EOT auf f_T wesentlich weniger stark aus. In Abbildung 5.13 ist zu sehen, dass eine Verkleinerung von $EOT = 2\text{nm}$ auf 0.5nm lediglich eine Erhöhung der maximalen f_T um 8% bewirkt. Die geringe Wirkung einer EOT -Verkleinerung kommt daher, dass dadurch sowohl g_m als auch C_{GG} in nahezu gleichem Ausmaß größer werden. Dies gilt jedoch nur, wenn externe parasitäre Kapazitäten, die unabhängig von EOT sind, vernachlässigt werden können. Da hier nur der innere Transistor betrachtet wird, ist das der Fall. Sind parasitäre Kapazitäten relevant, so ist eine deutlich stärkere EOT -Abhängigkeit von f_T zu erwarten. Dieser Fall wird in Abschnitt 5.3.2 ausführlich diskutiert.



(a)



(b)

Abbildung 5.12 Einfluss von EOT auf die Gleichstromeigenschaften von DG MOSFETs mit 10nm Gatelänge. (a) Transferkennlinien im linearen und logarithmischen Maßstab. (b) Maximal möglicher On-Strom bei gleichzeitiger Erfüllung des I_{off} -Kriteriums als Funktion von EOT .

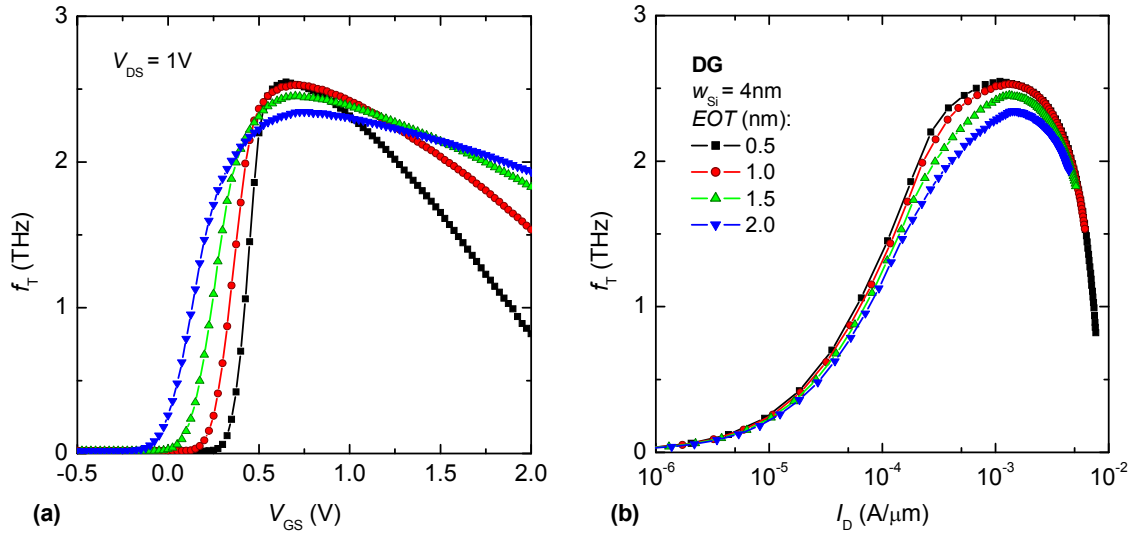


Abbildung 5.13 f_T von DG MOSFETs mit $L_G = 10\text{nm}$ und verschiedenen EOT als Funktion (a) der Gatespannung und (b) des Gatestroms.

Der Grund für die etwas geringere EOT -Abhängigkeit von C_{GG} im Vergleich zu der von g_m lässt sich folgendermaßen erklären. Im eingeschalteten Zustand ist sowohl für C_{GG} als auch für g_m die Reihenschaltung von Oxidkapazität und Inversionskapazität [Gleichung (3-37)] maßgeblich. Für g_m ist jedoch nur die flächenbezogene Gatekapazität am Sourceende des Kanals (genauer: bei x_m , der Stelle des Maximums der Potentialbarriere zwischen Source und Drain) von Bedeutung, während C_{GG} die integrale Gatekapazität ist. Man kann C_{GG} schreiben als

$$C_{GG} = \int_0^{L_G} C'_{GG}(x) dx, \quad (5-4)$$

wobei $C'_{GG}(x)$ die flächenbezogene Gatekapazität an der Stelle x entlang des Kanals ist. Diese ist gegeben durch

$$C'_{GG}(x) = \left(\frac{EOT}{\epsilon_{ox}} + \frac{y_{av}(x)}{\epsilon_{Si}} + \frac{1}{C_q} \right)^{-1} \quad (5-5)$$

und hängt vom mittleren Abstand der Kanalelektronen $y_{av}(x)$ von der nächstgelegenen Silizium/Gateisolator-Grenzfläche ab [$C_{es,Si}$ in Gleichung (3-37)].

In Abbildung 5.14 ist deutlich zu sehen, dass y_{av} am Sourceende des Kanals (bei x_m) besonders klein ist, in Richtung Drain größer wird und am Drainende des Kanals ein

Maximum hat. Am Sourceende des Kanals liegt der Ladungsschwerpunkt also viel näher an der Oberfläche als am Drainende. Nach Gleichung (5-5) ist aber der Einfluss von EOT auf C'_{GG} umso geringer, je größer $y_{av}(x)$ ist. Die integrale Gatekapazität C_{GG} kann also nicht so stark von EOT abhängen wie $C'_{GG}(x_m)$ und somit auch nicht so stark wie g_m .

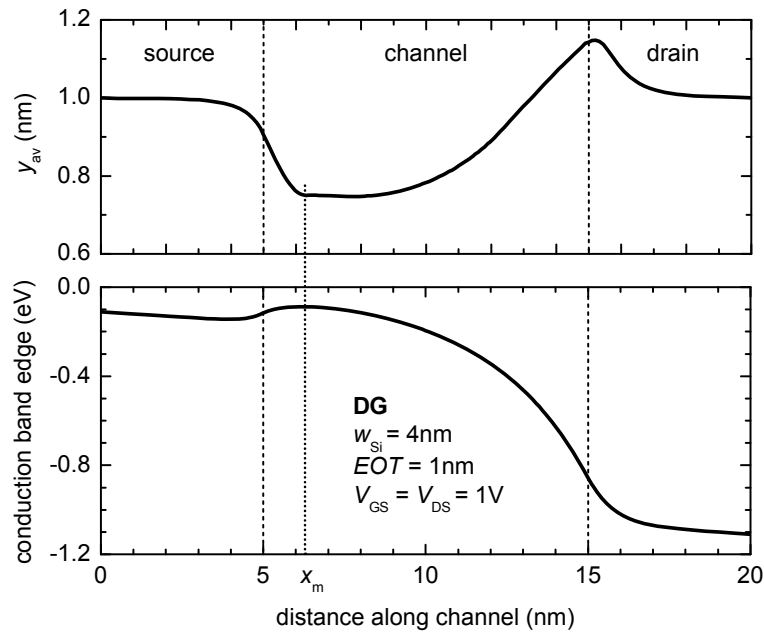


Abbildung 5.14 Mittlerer Abstand der Elektronen von der Silizium/Gateisolator-Grenzfläche (oben) und Leitbandbandkante (unten) entlang des Kanals eines DG MOSFET im On-Zustand.

In Abbildung 5.13(a) ist deutlich zu erkennen, dass in Richtung großer Gatespannungen f_T umso schneller mit V_{GS} abfällt, je kleiner EOT ist. Die Ursache dafür ist die stärkere Wirkung der Serienwiderstände bei kleinen EOT aufgrund der höheren Drainströme im Vergleich zu denen bei größeren EOT . Dadurch ist die Arbeitspunktverschiebung am inneren Transistor bei kleinen EOT besonders stark [siehe Gleichungen (4-15) und (4-16)]. Verschiebt sich der Arbeitspunkt in den linearen Bereich der Ausgangskennlinie, kommt es zu einer starken Vergrößerung von g_{DS} bei gleichzeitiger Verringerung von g_m des inneren Transistors. Entsprechend Gleichung (2-20) führt das zu dem beobachteten starken Abfall von f_T mit zunehmender Gatespannung, besonders bei kleinen EOT .

Betrachtet man f_T als Funktion des Drainstroms [Abb. 5.12(b)], so zeigt sich ein umgekehrtes Bild. Ein kleines EOT hat ein hohes f_T über einen größeren Bereich des

Drainstroms zur Folge. Insgesamt sind aber die Unterschiede zwischen den verschiedenen EOT -Varianten wesentlich geringer als die bei einer Variation der Siliziumdicke (vgl. Abschnitt 5.2.1).

In Abbildung 5.15 wird deutlich, dass f_{\max} stärker von EOT abhängt als f_T . Bei einer Verringerung der äquivalenten Oxiddicke von 2nm auf 0.5nm vergrößert sich f_{\max} um etwa 20%, während f_T nur um 8% größer wird. Das lässt sich mit Hilfe der analytischen f_{\max} -Formel, Gleichung (2-21), gut verstehen. Der g_{DS} -Term im Nenner von (2-21) spielt dabei die entscheidende Rolle. Mit Verkleinerung von EOT werden Kurzkanaleffekte zunehmend besser unterdrückt, wodurch auch der Drainleitwert kleiner wird. Um genau zu sein, halbiert sich g_{DS} im betrachteten EOT -Bereich. Dadurch wird der Nenner von (2-21) kleiner, was die im Vergleich zu f_T stärkere Zunahme von f_{\max} erklärt.

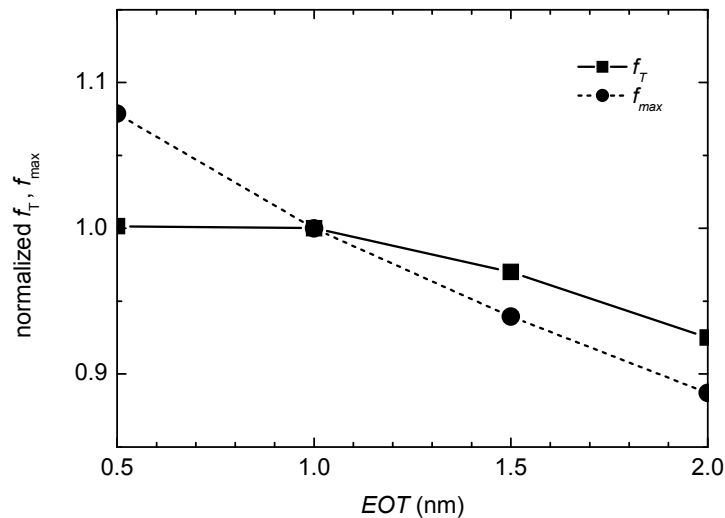


Abbildung 5.15 Maximale f_T und f_{\max} von DG MOSFETs mit $w_{Si} = 4\text{nm}$ als Funktion der äquivalenten Oxiddicke.

Abbildung 5.16 zeigt den Einfluss von EOT auf den Drainstrom bei $f_T = 50\text{GHz}$. Verglichen mit der Wirkung der Si-Dicke ist der Einfluss von EOT auf $I_D(50\text{GHz})$ relativ gering. Bei $w_{Si} = 4\text{nm}$ bewirkt eine Vergrößerung von w_{Si} um 1.5nm eine Erhöhung von $I_D(50\text{GHz})$ um etwa 35%, während eine EOT -Vergrößerung von 0.5nm auf 2nm $I_D(50\text{GHz})$ um etwa 17% erhöht. Der Grund für die geringere Wirkung von EOT liegt in dessen gegenläufigen Einflüssen auf S und C_{GG} (Abb. 5.17). Während S aufgrund zunehmender Kurzkanaleffekte mit EOT größer wird, nimmt C_{GG} mit

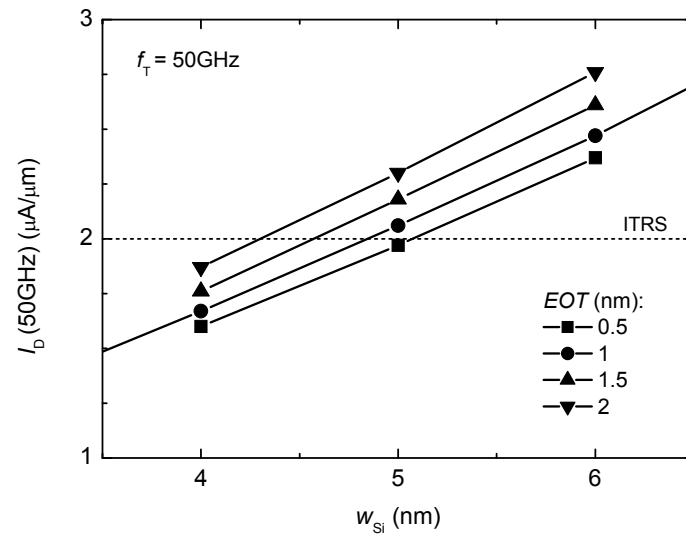


Abbildung 5.16 Drainstrom bei $f_T = 50\text{GHz}$ als Funktion der Si-Dicke mit EOT als Parameter.

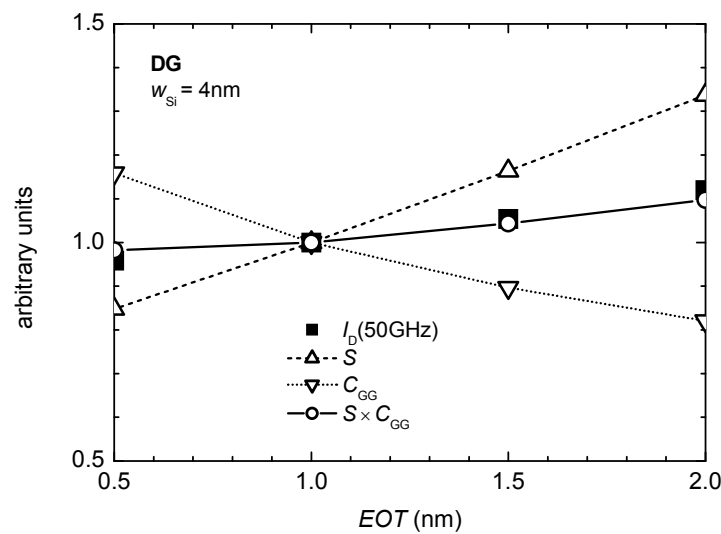


Abbildung 5.17 I_D , S und C_{GG} bei $f_T = 50\text{GHz}$ sowie das Produkt $S \times C_{GG}$ als Funktion von EOT .

wachsendem EOT ab. Beide Trends kompensieren sich zum größten Teil entsprechend Gleichung (5-3). Letztlich überwiegt der Einfluss von S , d.h. der Einfluss der Kurzkanaleffekte, so dass $I_D(50\text{GHz})$ größer wird.

Eine bemerkenswerte Schlussfolgerung aus dieser Untersuchung ist, dass der Einfluss von Kurzkanaleffekten auf $I_D(50\text{GHz})$ umso größer wird, je größer der Anteil parasitärer, von EOT unabhängiger Kapazitäten an C_{GG} ist. Sind solche Kapazitäten

nicht vernachlässigbar, dann ist die Variation des Drainstroms im hier betrachteten EOT -Bereich größer als nur 17%. Im Extremfall, d.h. bei totaler Dominanz parasitärer Kapazitäten wird C_{GG} unabhängig von EOT und $I_D(50\text{GHz})$ hängt genauso stark von EOT ab wie S . In diesem Fall würde eine Änderung von EOT von 0.5nm auf 2nm eine Erhöhung des Drainstroms um 58% bewirken.

5.3. Einfluss parasitärer Elemente

Nachdem der Einfluss der Transistorgeometrie auf die Hochfrequenzeigenschaften des inneren Transistors analysiert wurde, kommen wir nun zur Wirkung externer parasitärer Elemente. Im Folgenden wird zunächst die Wirkung parasitärer Widerstände auf die Grenzfrequenzen f_T und f_{max} untersucht und danach der Einfluss parasitärer Kapazitäten. Abschließend wird versucht, den Spielraum zulässiger Kombinationen von Widerständen und Kapazitäten einzugrenzen, mit denen ein Erreichen der Vorgaben für f_T und f_{max} möglich ist.

5.3.1. Parasitäre Widerstände

Abbildung 5.18 zeigt die Auswirkung externer Gate-, Source- und Drainwiderstände auf die Grenzfrequenzen f_T und f_{max} . Die dargestellten Ergebnisse wurden mit drei verschiedenen Methoden bestimmt, nämlich mit

- (I) ATLAS Simulationen, in denen die Widerstände direkt berücksichtigt wurden;
- (II) einem Kleinsignalersatzschaltbild (KSESB) bestehend aus den mit ATLAS bestimmten Y-Parametern des inneren Transistors und externen Widerständen;
- (III) Gleichungen (2-20) und (2-21), wobei die Parameter des inneren Transistors mit ATLAS bestimmt wurden.

Bei Methode (I) wurde die Kleinsignalanalyse (S^3A) an einer DG MOSFET Struktur mit kompakten Widerständen an den Gate-, Source- bzw. Drainkontakten durchgeführt. Dabei wurde sowohl die gleichstrommäßige als auch die kleinsignalmäßige Wirkung der Widerstände berücksichtigt. Im Unterschied dazu bleibt bei den Methoden (II) und

(III) eine Verschiebung des Gleichstromarbeitspunktes [vgl. (4-15) und (4-16)] unberücksichtigt.

Methode (II) funktioniert folgendermaßen. Nach den Regeln der Vierpoltheorie [155] werden die Y-Parameter des Gesamtvierpols aus innerem Transistor und externen Widerständen berechnet [156]. Mit Hilfe von (2-11) und (2-12) lassen sich daraus dann die Kurzschlussstromverstärkung bzw. die unilaterale Leistungsverstärkung für die betrachtete Frequenz bestimmen und schließlich entsprechend (2-15) und (2-16) f_T und f_{\max} extrapolieren.

Betrachten wir zunächst die Wirkung der Widerstände auf f_T [Abb. 5.18(a)]. In Übereinstimmung mit Gleichung (2-20) ist f_T völlig unabhängig vom Gatewiderstand R_G . Demgegenüber haben R_S und R_D vor allem durch ihre kleinsignalmäßige Wirkung eine deutliche Verringerung von f_T zur Folge, welche mit Gleichung (2-20) sehr gut beschrieben wird. Die Verschiebung des Gleichstromarbeitspunktes bewirkt eine zusätzliche Verringerung von f_T , welche aber erst bei sehr großen Widerständen von Bedeutung ist. So ist selbst bei $R_S = R_D = 500\Omega\mu\text{m}$ das mit Methode (I) bestimmte f_T , d.h. mit Berücksichtigung der Arbeitspunktverschiebung, nur um etwa 6% geringer als das f_T aus der rein kleinsignalmäßigen Betrachtung von Methode (II). Insgesamt trägt die Arbeitspunktverschiebung weniger als 3% zum gesamten, durch externe Widerstände verursachten f_T -Abfall bei.

Im Gegensatz zu f_T zeigt f_{\max} eine sehr starke Abhängigkeit von R_G , welche sehr gut durch Gleichung (2-21) beschrieben wird. Im betrachteten Bereich bis $500\Omega\mu\text{m}$ ist die Wirkung des Gatewiderstands sogar stärker als die gleich großer R_S und R_D zusammengenommen. Das lässt sich gut mit Gleichung (2-21) verstehen: solange die Widerstände klein sind, gehen in den g_{DS} -Term des Nenners von (2-21) R_G und R_S mit ähnlicher Wichtung ein, während im anderen Term die Wirkung von R_D gegenüber der von R_G um den Faktor $C_{GD}/(C_{GS} + C_{GD})$ (~ 0.2) abgeschwächt wird. Mit größer werdenden R_S und R_D wird jedoch f_T entsprechend (2-20) zunehmend kleiner und die Funktion N größer, was insgesamt zu einer Verstärkung der Wirkung von R_S und R_D gegenüber der von R_G führt. Für Widerstände jenseits von $500\Omega\mu\text{m}$ wird dann die Wirkung von R_S und R_D größer als die von R_G . Generell zeigt Abbildung 5.18, dass f_{\max} wesentlich stärker von parasitären Widerständen beeinflusst wird als f_T .

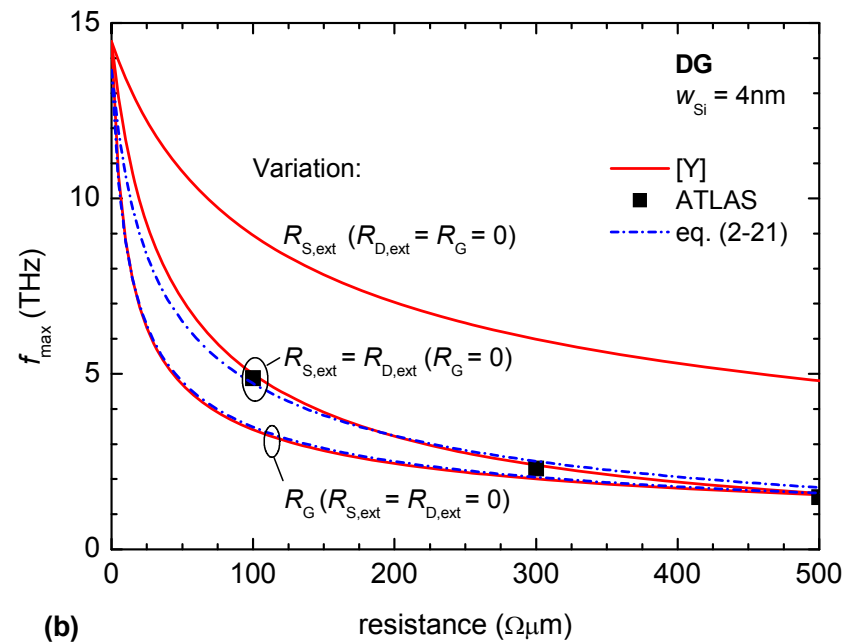
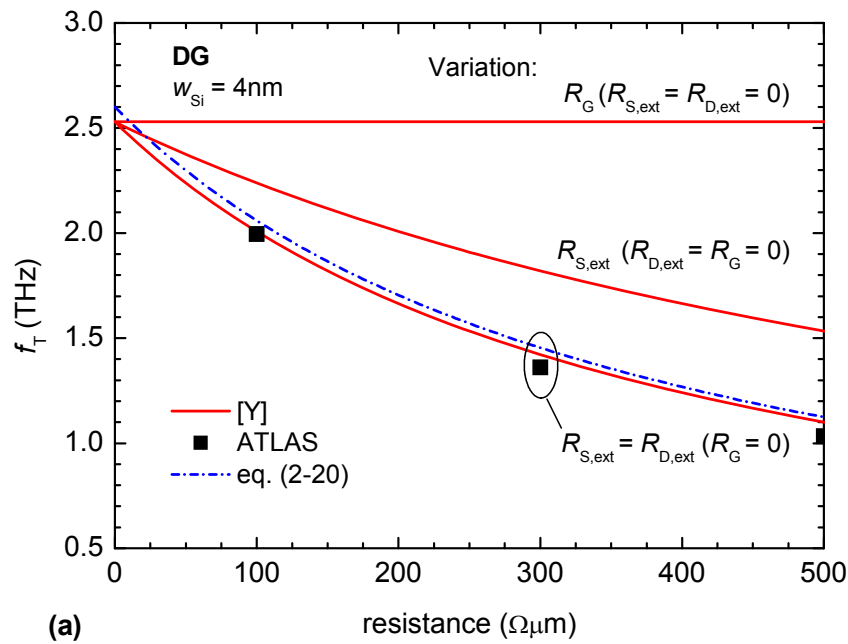


Abbildung 5.18 (a) f_T und (b) f_{max} in Abhängigkeit von externen parasitären Widerständen. Linien: Berechnungen nach KSESB (durchgezogen) oder mit Gleichungen (2-20) bzw. (2-21) (Strich-Punkt), in beiden Fällen ohne Berücksichtigung der Arbeitspunktverschiebung. Symbole: ATLAS-Simulationen mit kompakten Widerständen an Gate-, Source- und Drainkontakten, d.h. mit Berücksichtigung der Arbeitspunktverschiebung.

Bedeutung des Transistordesigns für die Wirkung parasitärer Widerstände

Da die Wirkung von parasitären Widerständen auf f_T und f_{\max} derart groß ist, stellt sich die Frage, inwiefern die Empfindlichkeit eines Transistors gegenüber parasitären Widerständen von seinem Design abhängt. Eine Antwort darauf lässt sich aus Abbildung 5.19 entnehmen. Sie zeigt f_T als Funktion der Si-Dicke von DG MOSFETs mit und ohne externe Widerstände von $100\Omega\mu\text{m}$ an den Kontakten. Während ohne externe Widerstände, wie bereits in Abschnitt 5.2.1 diskutiert, f_T mit zunehmender Schichtdicke leicht ansteigt, ist mit externen Widerständen ein deutliches Absinken von f_T mit zunehmender Schichtdicke festzustellen. Der entscheidende Grund dafür ist die starke Zunahme von g_{DS} bei einer Vergrößerung von w_{Si} , wie in Abb 5.19(b) deutlich zu sehen ist. Dadurch wird entsprechend Gleichung (2-20) die Wirkung von R_S und R_D verstärkt. Das Fazit lautet also: je besser ein Transistordesign Kurzkanaleffekte⁴ unterdrückt, desto unempfindlicher reagiert der Transistor auf parasitäre Widerstände.

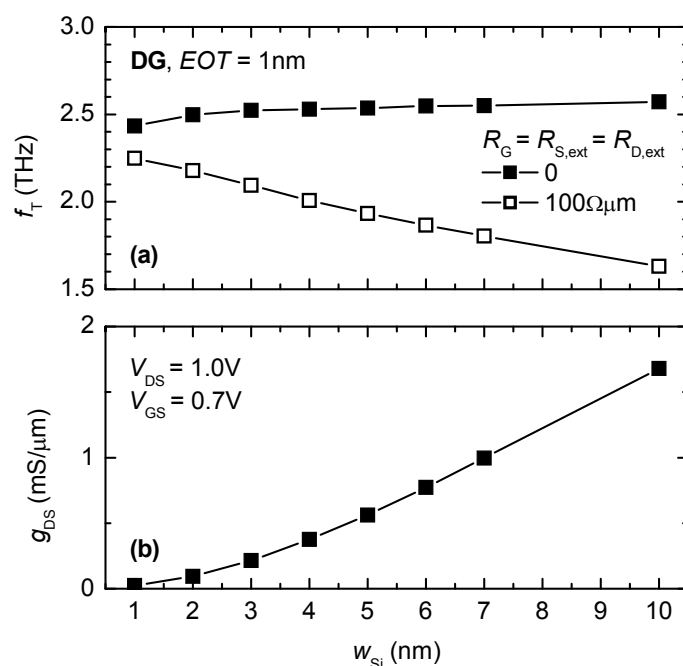


Abbildung 5.19 (a) f_T als Funktion der Siliziumdicke von DG MOSFET mit und ohne externe Widerstände von $100\Omega\mu\text{m}$ an Gate, Source und Drain. Die Wirkung externer Widerstände wurde mit Hilfe des KSESB [Methode (II)] berechnet. (b) Ausgangsleitwert des inneren Transistors als Funktion von w_{Si} .

⁴ Der Drainleitwert g_{DS} ist ein Maß für Kurzkanaleffekte: je größer $DIBL$, desto größer ist auch g_{DS} .

Abbildung 5.20 zeigt für DG MOSFETs mit $w_{\text{Si}} = 4\text{nm}$ und $w_{\text{Si}} = 6\text{nm}$ f_{T} und f_{max} als Funktion parasitärer Widerstände an Gate, Source und Drain. Ebenfalls eingezeichnet sind die Zielgrößen für f_{T} und f_{max} . An den Schnittpunkten der f_{T} - und f_{max} -Kurven mit den entsprechenden Targets lassen sich die maximalen Widerstände ablesen, die gerade noch das Erreichen der Zielgrößen zulassen. Es zeigt sich, dass zum Erreichen des f_{max} -Targets generell kleinere Widerstände erforderlich sind als zum Erreichen des f_{T} -Targets, wobei die Widerstände am Transistor mit $w_{\text{Si}} = 6\text{nm}$ jeweils deutlich kleiner sein müssen als am Transistor mit $w_{\text{Si}} = 4\text{nm}$. Zum Vergleich: der DG MOSFET mit $w_{\text{Si}} = 4\text{nm}$ toleriert Widerstände bis etwa $340\Omega\mu\text{m}$, während es bei $w_{\text{Si}} = 6\text{nm}$ höchstens $240\Omega\mu\text{m}$ sein dürfen, also rund 30% weniger. Diese Widerstandswerte sind jedoch nur als obere Grenzwerte anzusehen, da im realen Transistor auch noch parasitäre Kapazitäten wirksam werden, die hier noch nicht berücksichtigt wurden. Die Wirkung parasitärer Kapazitäten ist Gegenstand des nächsten Abschnitts.

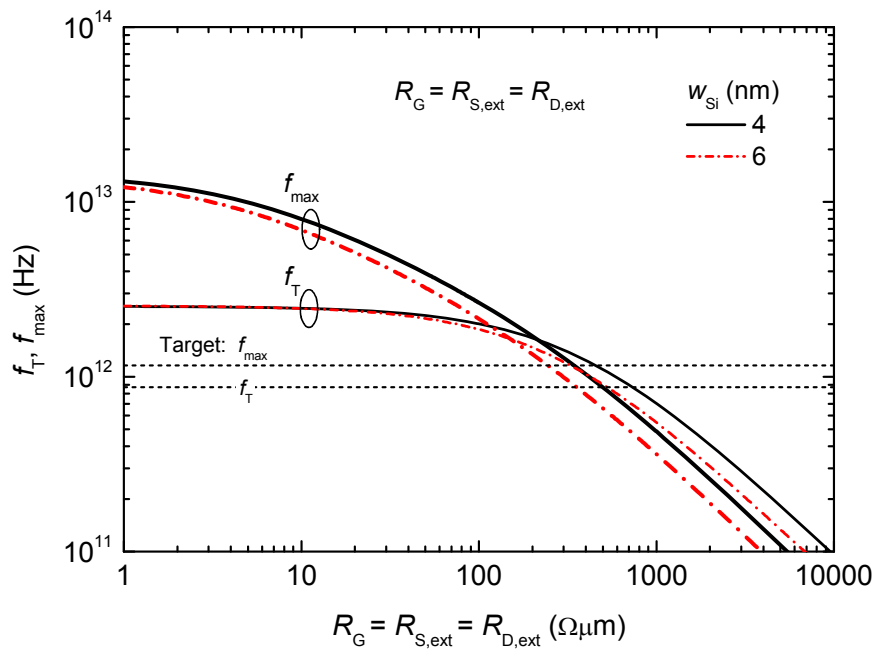


Abbildung 5.20 f_{T} und f_{max} von DG MOSFETs mit $w_{\text{Si}} = 4\text{nm}$ und $w_{\text{Si}} = 6\text{nm}$ als Funktion parasitärer Widerstände. Horizontale Linien: Zielgrößen für f_{T} und f_{max} .

5.3.2. Parasitäre Kapazitäten

Neben parasitären Widerständen wirken sich vor allem parasitäre Kapazitäten negativ auf die Hochfrequenzeigenschaften stark skaliert MOSFETs aus. In den bisher untersuchten MOSFET-Strukturen sind solche unerwünschten Kapazitäten, die nicht für die Funktion des Transistors benötigt werden, auf ein Minimum begrenzt. Im Vergleich dazu kann in realen MOSFETs die dreidimensionale Anordnung von Gateelektrode und Source/Drain-Anschlussgebieten zu einer erheblichen Vergrößerung von C_{GS} und C_{GD} führen. Die Wirkung solcher parasitärer Kapazitäten auf das Kleinsignalverhalten wurde mit Hilfe des KSESB [Methode (II)] berechnet.

Abbildung 5.21 zeigt für einen DG MOSFET mit Gate-, Source- und Drainwiderständen von jeweils $75\Omega\mu\text{m}^{(5)}$ f_T und f_{\max} in Abhängigkeit parasitärer C_{GS} und C_{GD} , im weiteren mit $C_{GS,\text{par}}$ und $C_{GD,\text{par}}$ bezeichnet. Für $C_{GS,\text{par}}$ und $C_{GD,\text{par}} > 10^{-17}$ F/ μm ($= 10^{-2}$ fF/ μm), das sind jeweils etwa 3% der Oxidkapazität des inneren Transistors, ist eine merkliche Verkleinerung von f_T und f_{\max} festzustellen. Bei $C_{GS,\text{par}}$ und $C_{GD,\text{par}} > 0.16$ fF/ μm wird f_{\max} kleiner als die entsprechende Zielgröße, bei etwas größeren Kapazitäten (0.175fF/ μm) wird auch das Target für f_T nicht mehr erreicht. Der betrachtete DG MOSFET erreicht also die Vorgaben nur dann, wenn die Summe der parasitären Kapazitäten höchstens 90% der Oxidkapazität beträgt.

In Abbildung 5.21 wird außerdem deutlich, dass insbesondere f_{\max} , aber auch f_T , wesentlich stärker von C_{GD} abhängt als von C_{GS} . Das lässt sich sehr gut mit Hilfe von (2-20) und (2-21) verstehen. Im Nenner von (2-20) wird über das Verhältnis $C_{GD}/(C_{GS} + C_{GD})$ die Wirkung von C_{GD} auf f_T gegenüber der von C_{GS} verstärkt. Aus Gleichung (2-20) folgt aber auch, dass sich die Wirkungen von C_{GS} und C_{GD} auf f_T umso weniger unterscheiden, je kleiner die Serienwiderstände R_S und R_D sind. Tatsächlich liefern die Berechnungen nach dem KSESB ohne externe Widerstände identische Wirkungen von C_{GS} und C_{GD} auf f_T .

Bei Gleichung (2-21) ist der zweite Term des Nenners entscheidend für den unterschiedlich starken Einfluss von C_{GS} und C_{GD} auf f_{\max} . Dieser Term ist direkt proportional zu C_{GD} . Dadurch wird f_{\max} viel stärker durch C_{GD} reduziert als durch ein gleich großes C_{GS} .

⁵ Die ITRS fordert für LSTP-Logik $R_S + R_D \leq 160\Omega\mu\text{m}$ [4].

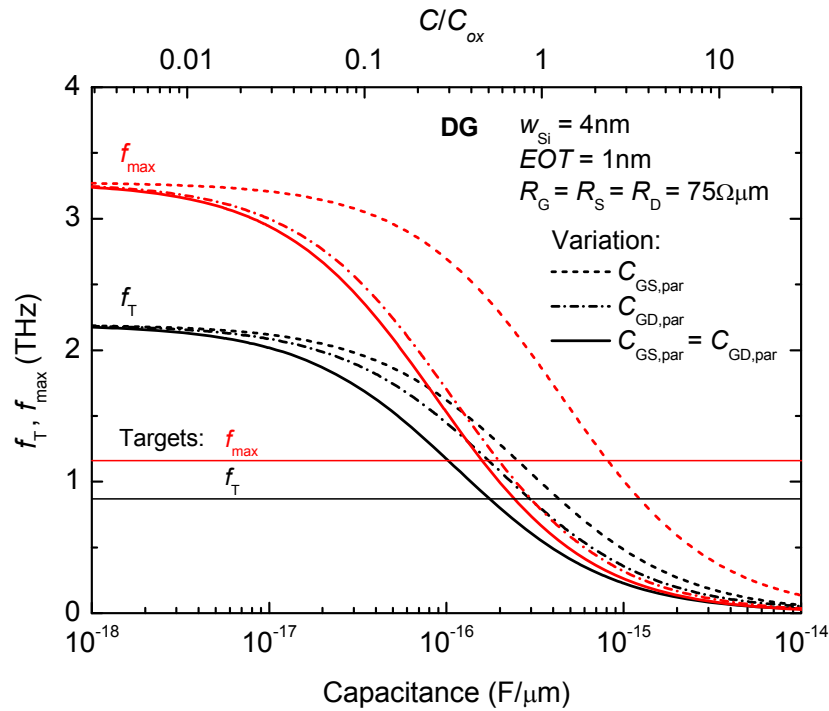


Abbildung 5.21 f_T und f_{max} eines DG MOSFETs mit $w_{Si} = 4\text{nm}$ in Abhängigkeit von parasitären Kapazitäten. Horizontale Linien: Zielgrößen für f_T und f_{max} .

Die Wirkung parasitärer Kapazitäten in Abhängigkeit vom Transistordesign

Auch der Einfluss parasitärer Kapazitäten auf f_T hängt vom Design des inneren Transistors ab, insbesondere von der äquivalenten Oxiddicke. Das lässt sich sehr gut anhand von Gleichung (2-22) verstehen. Wie bereits in Abschnitt 5.2.2 diskutiert, hängen sowohl g_m als auch C_{GS} und C_{GD} des inneren Transistors ähnlich stark von EOT ab, so dass sich f_T des inneren Transistors nur schwach mit EOT ändert. Kommen nun aber noch parasitäre Kapazitäten ($C_{GS,par}$ bzw. $C_{GD,par}$) hinzu, die nicht oder nur schwach von EOT abhängen, so verringert sich die EOT -Abhängigkeit des Nenners von (2-22) gegenüber der des Zählers (g_m). In dem Fall muss f_T wesentlich stärker von EOT abhängen als ohne parasitäre Kapazitäten. In Abbildung 5.22 ist dieser Effekt deutlich erkennbar. Mit $C_{GS,par} + C_{GD,par} = 0.1\text{fF}/\mu\text{m}$ liefern die Berechnungen nach dem KSESb für eine Vergrößerung von EOT von 0.5 auf 2nm eine Verringerung von f_T um 27%, während sich f_T ohne parasitäre Kapazitäten nur um 8% verringert.

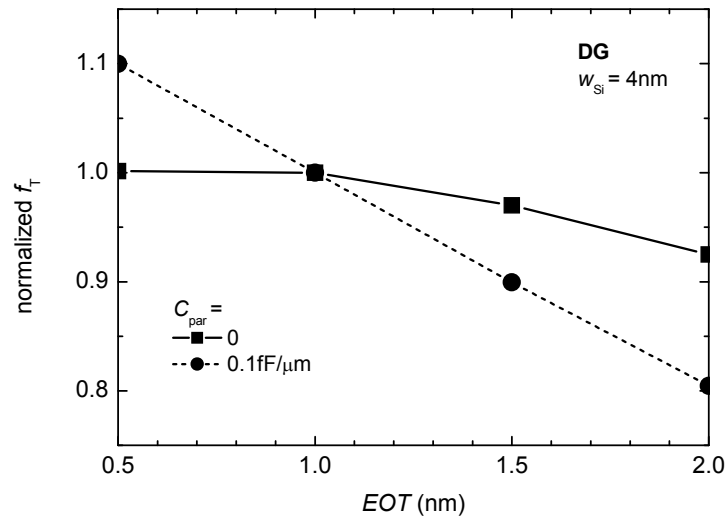


Abbildung 5.22 f_T eines DG MOSFETs mit und ohne parasitäre Kapazitäten in Abhängigkeit von der äquivalenten Oxiddicke.

Zulässige Kombinationen von parasitären Widerständen und Kapazitäten

Da die Grenzfrequenzen eines Transistors sowohl durch parasitäre Kapazitäten als auch durch parasitäre Widerstände reduziert werden, ist es wichtig zu wissen, mit welchen Kombinationen von Widerständen und Kapazitäten die Vorgaben für f_T und f_{max} erreichbar sind. Klar ist, je größer die Widerstände R_G , R_S und R_D eines MOSFET sind, desto kleiner müssen seine parasitären Kapazitäten sein, wenn er die Targets für f_T und f_{max} erreichen soll und umgekehrt. Die genauen Grenzwerte für bestimmte R/C -Kombinationen hängen außerdem vom Design des inneren Transistors ab, insbesondere von der Siliziumdicke und der äquivalenten Oxiddicke. In Abbildung 5.23 sind die Grenzwerte erlaubter R/C -Kombinationen für drei verschiedene DG MOSFET-Designs dargestellt. Für diese Darstellung wurde $R_G = R_S = R_D$ sowie $C_{GS,par} = C_{GD,par}$ angenommen. Eine Kurve in Abbildung 5.23 gibt die maximal erlaubte parasitäre Gatekapazität $C_{par} = C_{GS,par} + C_{GD,par}$ als Funktion von R_S an, bei der eine bzw. beide Zielgrößen gerade noch erreicht werden. Eine gestrichelte Kurve begrenzt den Bereich zum Erreichen des f_T -Targets, eine strich-punktierte Kurve den Bereich zum Erreichen des f_{max} -Targets für ein bestimmtes MOSFET-Design. Für alle R/C -Kombinationen unterhalb bzw. links einer durchgezogenen Kurve ist sowohl f_T als auch f_{max} des betrachteten MOSFET größer als die entsprechende Zielgröße.

Die durchgezogenen Kurven zeigen klar an, dass für alle Designs bei kleinen Widerständen das zulässige C_{par} durch das f_{T} -Kriterium begrenzt wird, während bei kleinen Kapazitäten die erlaubten Widerstände vom f_{max} -Kriterium begrenzt werden. Es wird außerdem deutlich, dass eine Vergrößerung von EOT vor allem die maximal erlaubten C_{par} reduziert, während bei einer Vergrößerung der Siliziumdicke in erster Linie die maximal zulässigen Widerstände kleiner werden. Der Vergleich der drei MOSFET-Designs in Abbildung 5.23 macht deutlich, dass das Design mit den besten Gleichstromeigenschaften ($w_{\text{Si}} = 4\text{nm}$, $EOT = 1\text{nm}$), d.h. das Design das Kurzkanaleffekte am besten unterdrückt und den größten On-Strom liefern kann (vgl. Abb. 5.2 und 5.12), den größten Spielraum für parasitäre Kapazitäten und Widerstände zulässt.

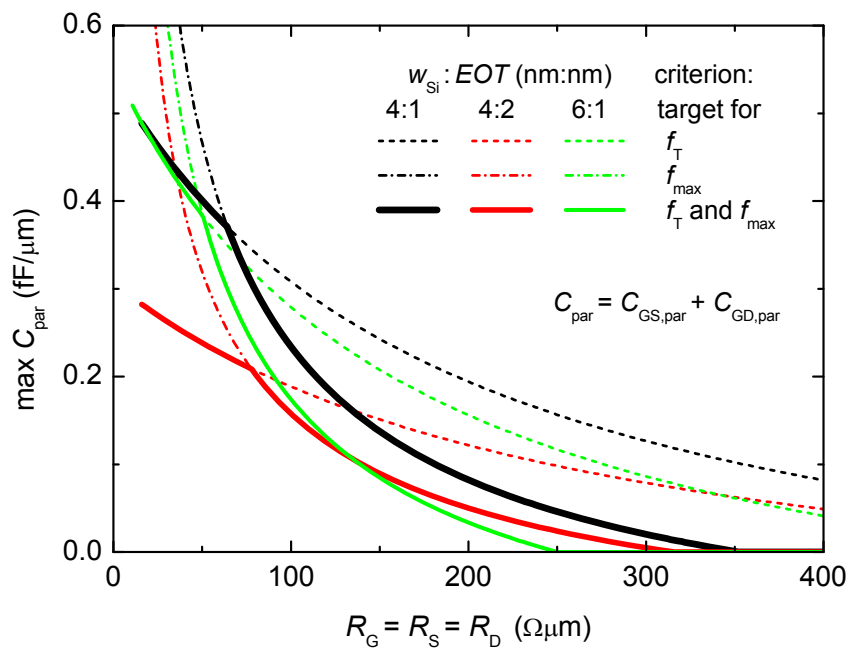


Abbildung 5.23 Obergrenzen zulässiger Kombinationen von parasitären Widerständen und Kapazitäten für drei verschiedene DG MOSFET-Designs, die ein Erreichen der Zielgrößen für f_{T} (gestrichelte Linien), f_{max} (Strich-Punkt-Linien) sowie für f_{T} und f_{max} gleichzeitig (durchgezogene Linien) ermöglichen.

Fazit

Kommen wir noch einmal auf die beiden eingangs gestellten Fragen zurück. Sie lauteten

- (1) Gelten für optimale Hochfrequenzeigenschaften ähnliche Designregeln wie für optimale Gleichstromeigenschaften? Oder: *Muss ein guter HF MOSFET auch ein guter Digital-MOSFET sein?*
- (2) Unter welchen Voraussetzungen sind die Vorgaben der ITRS für f_T und f_{\max} erreichbar?

Abbildung 5.23 gibt eine Antwort auf Frage (2). Für ein Erreichen der extrem hohen Vorgaben für f_T und f_{\max} sind der Größe von parasitären Widerständen und Kapazitäten enge Grenzen gesetzt. Als Grundregel lässt sich festhalten, dass die Summe parasitärer Kapazitäten deutlich kleiner sein muss als die intrinsische Gateoxidentkapazität. Einen guten Richtwertwert für die Source- und Drainwiderstände stellen die Vorgaben der ITRS für LSTP CMOS Logik dar ($\sim 80\Omega\mu\text{m}$). Der Gatewiderstand sollte nicht größer sein als R_S und R_D .

Auf die erste Frage liefern die vorangegangenen Untersuchungen folgende Antwort: *Ein guter Digital-MOSFET ist auch ein guter HF MOSFET*. Gute Gleichstromeigenschaften machen die Grenzfrequenzen des Transistors unempfindlicher gegenüber parasitären Widerständen und Kapazitäten. Darüber hinaus ist der für ein bestimmtes f_T notwendige Drainstrom umso geringer, je besser die Gleichstromeigenschaften des Transistors sind. Es ist also grundsätzlich richtig, sich beim Design von HF MOSFETs an den Vorgaben für CMOS Logik zu orientieren. Die Ergebnisse aus Abbildung 5.23 zeigen aber auch, dass hinsichtlich der Gleichstromeigenschaften weniger gut designte MOSFETs die Vorgaben für f_T und f_{\max} durchaus erfüllen können, wenn es gelingt deren parasitäre Widerstände und Kapazitäten entsprechend zu begrenzen.

Hinsichtlich der ITRS Vorgabe zum Drainstrom bei einem f_T von 50GHz ist das DG-Konzept dem SG SOI-Konzept eindeutig vorzuziehen (vgl. Abb. 5.6).

Kapitel 6

Zusammenfassung und Ausblick

Der Silizium-MOSFET ist nach wie vor das Standardbauelement der Halbleiterelektronik und wird es aller Voraussicht nach bis in absehbare Zukunft auch bleiben. Geht man davon aus, dass der Trend der letzten Jahrzehnte beibehalten wird (Stichwort: Moore'sches Gesetz), so ist um das Jahr 2020 mit der Massenproduktion von integrierten Schaltkreisen zu rechnen, deren Transistoren eine Gatelänge von nur noch 10nm besitzen. Bei so kleinen Abmessungen kann die Funktionsweise der Transistoren nur durch die Anwendung neuartiger bzw. nichtklassischer MOSFET-Konzepte aufrechterhalten werden. Darüber hinaus muss mit einer Reihe quantenmechanischer Effekte gerechnet werden, über deren tatsächliche Auswirkungen auf die Performance oder die Zuverlässigkeit der Transistoren noch keine endgültige Klarheit besteht.

Das Ziel der vorliegenden theoretischen Arbeit war es, verschiedene nichtklassische MOSFET-Konzepte hinsichtlich ihrer Performance sowie ihrer Eignung für zukünftige Logik- bzw. Analog/RF-Anwendungen zu bewerten und ihren Designspielraum einzugrenzen. Zu diesem Zweck war es zunächst erforderlich ein geeignetes Simulationsmodell zu finden, das es ermöglicht, eine Vielzahl unterschiedlicher Transistorstrukturen in einem weiten Designbereich mit angemessener physikalischer Genauigkeit zu beschreiben. Der Fokus lag auf MOSFETs mit einer Gatelänge von 10nm. In so kurzen Kanälen ist mit einem starken Einfluss nichtstationärer bzw. quasiballistischer Transporteffekte zu rechnen, deren korrekte Behandlung im Prinzip nur mit Hilfe von Monte-Carlo Simulationen möglich ist. Aufgrund ihres enormen rechentechnischen Aufwandes kam die Monte-Carlo Methode jedoch nicht als Standardverfahren zur Bearbeitung der gestellten Aufgabe infrage. Ein detaillierter Vergleich verschiedener Simulationsmodelle an ausgewählten Transistorstrukturen ergab, dass das einfache und robuste Drift-Diffusionsmodell mit einer modifizierten $v(E)$ -Charakteristik den Anforderungen am besten genügt. Es konnte gezeigt werden, dass damit nicht nur das Gleichstromverhalten von extrem skalierten MOSFETs hinreichend genau simuliert werden kann, sondern auch das Kleinsignal- und Hochfrequenzverhalten.

Ein weiterer Schwerpunkt lag darin, die Auswirkung von Quanteneffekten auf die elektrische Performance extrem skaliert MOSFETs abzuschätzen. Zu diesem Zweck wurden analytische Modelle entwickelt, mit denen sich die Auswirkungen des Quantum Confinement in Single-Gate SOI sowie in Multiple-Gate MOSFETs auf die Schwellspannung berechnen lassen. Eine weitere Studie beschäftigte sich ausführlich mit dem Einfluss von Quanteneffekten auf die Gatekapazität von Tri-Gate MOSFETs. Darüber hinaus wurden die Auswirkungen der Subbandformierung auf die Elektronenbeweglichkeit in Tri-Gate MOSFETs untersucht. Schließlich wurden einfache Möglichkeiten demonstriert, mit deren Hilfe sich klassisch simulierte Kennlinien nachträglich so korrigieren lassen, dass die Auswirkungen von Quanteneffekten berücksichtigt werden können.

Im Rahmen dieser Arbeit wurden zwei Studien angefertigt, die sich mit dem Design und der Performance von nichtklassischen MOSFETs mit einer Gatelänge von 10nm befassen. Eine Studie konzentrierte sich auf MOSFETs für digitale High-Performance Logik, die andere auf MOSFETs für Analog/RF-Anwendungen. Die Studie zu den Digital-MOSFETs lieferte im Wesentlichen folgende Ergebnisse:

- Der Designspielraum für die Querschnittsgeometrie von Tri-Gate (TG) MOSFETs, inklusive der beiden Spezialfälle Single-Gate (SG) und Double-Gate (DG) MOSFET, konnte eingegrenzt werden.

Dieser Spielraum ist zu großen Kanalquerschnitten hin durch zunehmende Kurzkanaleffekte begrenzt. Diese haben zur Folge, dass die Kriterien für On- und Off-Strom nicht mehr gleichzeitig erfüllt werden können.

Die untere Grenze des Designspielraums wird durch die zunehmenden Auswirkungen des Quantum Confinement bestimmt. Diese heben eine enorme Zunahme der Empfindlichkeit des Off-Stroms gegenüber Schwankungen der Querschnittsgeometrie zur Folge. Bei Designs unterhalb dieser unteren Grenze haben bereits Schwankungen der Abmessungen um nur eine Atomlage Fluktuationen des Off-Stroms um mehr als den Faktor zwei zur Folge, und zwar nur aufgrund von Quanteneffekten. Dazu kommen dann noch die Auswirkungen der klassischen Elektrostatik (Kurzkanaleffekte usw.).

- Aus den Grenzen des Designspielraums ergibt sich, dass sowohl SG, DG als auch TG MOSFETs für High-Performance Logikanwendungen prinzipiell infrage kommen.

- Der Einfluss unvermeidbarer Geometrietoleranzen auf die elektrischen Parameter On-Strom, Off-Strom und Schwellspannung von SG, DG und TG MOSFETs wurde ausführlich untersucht.

Diese Untersuchungen haben gezeigt, dass die Massenfertigung von MOSFETs mit $L_G = 10\text{nm}$ extreme Anforderungen an die Prozesskontrolle stellt, um Fluktuationen vor allem des Off-Stroms in vertretbaren Grenzen zu halten. Die Toleranzen der Kanalgeometrie sollten im Schnitt deutlich unter 10% liegen, d.h. eher im Bereich der Dicke einer Atomlage.

- Der Vergleich von Single-Gate, Double-Gate und Tri-Gate MOSFETs liefert kein bevorzugtes MOSFET-Design für diesen Gatelängenbereich. Diese Aussage weicht von den Ergebnissen anderer Studien ab, die das DG-Konzept [146], [149]-[150] gegenüber dem TG-Konzept favorisieren. Dagegen ist die Empfehlung der vorliegenden Studie, das Design so zu wählen, dass mit der zur Verfügung stehenden Technologie die Performance-Schwankungen möglichst gering sind. Dementsprechend hat auch das SG SOI-Konzept durchaus seine Berechtigung. Das Fazit lautet also: *Für jede Technologie gibt es ein optimales MOSFET-Design.*
- Die Auswirkungen parasitärer Widerstände auf die Gleichstromeigenschaften von MOSFETs mit 10nm Gatelänge, insbesondere auf den On-Strom wurden untersucht. Dabei wurde festgestellt, dass bereits eine Verdopplung der internen (minimalen) Bahnwiderstände der Simulationsstrukturen genügt, um das I_{on} -Kriterium zu verfehlen. Mit Serienwiderständen, die den Vorgaben der ITRS entsprechen, werden nur etwa 80% des I_{on} -Targets erreicht.
- Dieses Ergebnis unterstreicht die enorme Bedeutung der Minimierung parasitärer Widerstände in extrem skalierten MOSFETs. Zugleich wird deutlich, dass die Vorgaben der ITRS bezüglich On- und Off-Strom für High-Performance MOSFETs mit $L_G = 10\text{nm}$ möglicherweise nicht gleichzeitig erreichbar sind.

Die Ergebnisse der Studie zu MOSFETs für Analog/RF-Anwendungen lassen sich wie folgt zusammenfassen:

- Betrachtet man nur den inneren Transistor, d.h. Vernachlässigung externer parasitärer Widerstände und Kapazitäten, so zeigen alle MOSFET-Strukturen unabhängig vom Kanalquerschnitt extrem hohe f_T -Werte um 2.5THz. Das ist beinahe dreimal so hoch wie das ITRS-Target. Die simulierten f_{max} befinden sich

sogar im zweistelligen THz-Bereich, was im Wesentlichen auf den in den Simulationen vernachlässigten Gatewiderstand zurückzuführen ist.

Diese Ergebnisse zeigen, dass der Si MOSFET zumindest prinzipiell in der Lage ist, die ITRS-Targets zu erfüllen.

- Es zeigt sich, dass das f_T -Maximum bei umso kleineren I_D erreicht wird, je kleiner der Kanalquerschnitt ist. Das bedeutet, je besser die Gleichstromeigenschaften eines Transistors sind, desto geringer ist seine DC-Verlustleistung bei gleichem f_T .
- Für ein f_T von 50GHz genügt es, MOSFETs mit 10nm Gatelänge im Subthresholdbereich zu betreiben. Das verspricht für viele Gigahertzanwendungen eine extrem geringe DC-Verlustleistung.

Betrachtet man den Drainstrom, der für ein f_T von 50GHz benötigt wird, so ist das DG-Konzept dem SG SOI-Konzept eindeutig vorzuziehen.

- Der Auswirkungen parasitärer Widerstände und Kapazitäten auf die Grenzfrequenzen f_T und f_{\max} wurden ausführlich untersucht.
- Zum besseren Verständnis der Ergebnisse der numerischen Simulationen wurde ein analytisches Kompaktmodell für f_{\max} entwickelt, welches die Einflüsse parasitärer Elemente in einem weiten Bereich richtig beschreibt. Der Vorteil dieses Modells im Vergleich zu herkömmlichen Modellen für f_{\max} besteht darin, dass es auch die Wirkung des Drainwiderstands korrekt wiedergibt, welche üblicherweise unberücksichtigt bleibt.
- Es zeigt sich, dass f_{\max} wesentlich stärker von parasitären Widerständen beeinflusst wird als f_T .
- Ein DG MOSFET, dessen parasitäre Widerstände den Vorgaben der ITRS entsprechen, erreicht die Zielgrößen für f_T und f_{\max} nur dann, wenn die Summe der parasitären Kapazitäten höchstens 90% seiner intrinsischen Oxidkapazität beträgt.
- Für ausgewählte MOSFET-Designs wurde abschließend der Spielraum zulässiger Kombinationen von Widerständen und Kapazitäten eingegrenzt, mit denen die Vorgaben für f_T und f_{\max} noch erreicht werden können.

Der Vergleich zeigt, dass das Design mit den besten Gleichstromeigenschaften, d.h. das Design mit der besten Unterdrückung von Kurzkanaleffekten, den größten Spielraum für parasitäre Kapazitäten und Widerstände zulässt.

Insgesamt zeigen die Ergebnisse dieser Arbeit, dass der Silizium-MOSFET nach wie vor ein großes Potential für zukünftige Digital- und Hochfrequenzanwendungen besitzt. Für die Generation mit 10nm Gatelänge sind nichtklassische Multiple-Gate Konzepte besonders geeignet. Aber auch der Single-Gate SOI MOSFET ist prinzipiell in der Lage, die Anforderungen zu erfüllen.

Besonders beeindruckend sind die intrinsischen Hochfrequenzeigenschaften der untersuchten MOSFET-Strukturen. Wenn es gelingt, parasitäre Widerstände und Kapazitäten in akzeptablen Grenzen zu halten, so kommt der Silizium-MOSFET sogar für neue Anwendungen im Bereich des Terahertz-Gap (300GHz – 3THz) infrage.

Die berechneten hohen Grenzfrequenzen dieser stark skalierten MOSFETs im Subthresholdbereich legen noch ein weiteres Anwendungsgebiet nahe. Man könnte es mit *Ultra Low Power RF Electronics* bezeichnen. Ein großer Vorteil des Betriebs im Subthresholdbereich ist beispielsweise, dass parasitäre Widerstände einen wesentlich geringeren Einfluss auf die Grenzfrequenzen haben als im „normalen“ eingeschalteten Zustand. Erste Schritte zum Verständnis des Kleinsignalverhaltens im Subthresholdbereich wurden in der vorliegenden Arbeit aufgezeigt.

Ausblick

Wie jede wissenschaftliche Arbeit, so kann auch die vorliegende nur unvollständig sein. Grundsätzlich handelt es sich hier um eine rein theoretische Arbeit, die sich an experimentellen Ergebnissen messen lassen muss. Das war im Rahmen der Arbeit nur an wenigen Punkten indirekt möglich, ist aber für die Zukunft unvermeidlich.

Ein weiterer Schwerpunkt zukünftiger Arbeiten zu diesem Thema sollte auf der Vervollständigung und Verbesserung der Simulationsmodelle liegen. Insbesondere ein geeignetes Beweglichkeitsmodell, welches die Effekte des eindimensionalen Transports in Silizium-Nanodrähten richtig beschreibt ist in höchstem Maße wünschenswert. Ein derartiges Modell existiert derzeit weltweit noch nicht und muss vor allem die Geometrie- und Feldabhängigkeit der Beweglichkeit richtig wiedergeben.

Der Schwerpunkt der vorliegenden Arbeit liegt auf der Untersuchung der Eigenschaften und des Designs stark skalierten n-Kanal MOSFETs. Für zukünftige CMOS-Anwendungen sind aber p-Kanal MOSFETs ebenso relevant. Eine äquivalente Studie zum Design von pMOSFETs ist daher überaus wünschenswert. Voraussetzung dafür ist jedoch eine modifizierte $v(E)$ -Charakteristik für den Löchertransport in solchen Bauelementen, welche derzeit noch nicht existiert.

Anhang

A. Analytische Näherungslösungen für f_T und f_{\max}

Wendet man die Definitionsgleichungen der Y-Parameter (2-10) auf das Kleinsignalersatzschaltbild in Abbildung 2.9 an, so ergeben sich die Real- und Imaginärteile der Y-Parameter näherungsweise zu

$$Re(y_{11}) = (R_G + R_S)[Im(y_{11})]^2 - \frac{\omega^2 C_{GD}}{N^2} [R_S(C_{GS} + C_{GD}) + R_D g_m (C_{GS} R_S - C_{GD} R_D)] \quad (A-1)$$

$$Im(y_{11}) = \frac{\omega}{N} [(C_{GS} + C_{GD})N + g_m (C_{GD} R_D - C_{GS} R_S)] \quad (A-2)$$

$$Re(y_{12}) = -\frac{\omega^2 C_{GD}}{N^2} \left\{ R_G (C_{GS} + C_{GD}) \left[1 + \frac{C_{GD}}{C_{GS} + C_{GD}} g_m (R_S + R_D) \right] N + R_D C_{GD} N^2 + g_{DS} \left[(R_G g_m (R_S + R_D) + R_D N) (C_{GS} R_S - C_{GD} R_D) - N (C_{GD} R_D^2 + C_{GS} R_S^2) \right] \right\} \quad (A-3)$$

$$Im(y_{12}) = -\omega \left[C_{GD} + \frac{g_{DS}}{N} (C_{GS} R_S - C_{GD} R_D) \right] \quad (A-4)$$

$$Re(y_{21}) = \frac{g_m}{N} \quad (A-5)$$

$$Im(y_{21}) = -\frac{\omega}{N} [(C_{GS} + C_{GD}) g_m (R_G + R_S) + C_{GD}] - \frac{\omega g_m}{N^2} [g_m (R_G + R_S) (C_{GD} R_D - C_{GS} R_S) + C_{GD} R_D] \quad (A-6)$$

$$Re(y_{22}) = \frac{1}{N} (g_{DS} + \omega^2 C_{GD} R_G) \quad (A-7)$$

$$Im(y_{22}) = -\omega C_{GD} + \frac{\omega C_{GD}}{N} (g_m R_G - g_{DS} R_D) \quad (A-8)$$

mit

$$N = 1 + g_{DS} (R_S + R_D) + g_m R_S. \quad (A-9)$$

Gleichungen (A-1) – (A-8) sind gültig für Frequenzen deutlich unterhalb f_T unter der Voraussetzung, dass einerseits $C_{DS} \ll C_{GS}$, C_{GD} , und dass sich andererseits die Widerstände R_G , R_S und R_D nicht extrem voneinander unterscheiden. Die Forderung $f \ll f_T$ führt unter anderem dazu, dass der mit der Steilheit zusammenhängende Faktor $e^{-j\omega\tau}$ vernachlässigt werden kann, da τ in der Größenordnung von $1/2\pi f_T$ liegt. Darüber hinaus kann man von einem gut konstruierten MOSFET erwarten, dass im Sättigungsbereich

$$g_m(R_G + R_S) \leq 1, \quad g_m R_D \leq 1 \quad (\text{A-10})$$

sowie

$$g_{DS} \leq 0.1 g_m. \quad (\text{A-11})$$

Unter diesen Voraussetzungen liefert der Vergleich von (A-1) – (A-8)

$$|Re(y_{11})| \ll |Im(y_{11})| \ll |Re(y_{21})|, \quad (\text{A-12})$$

$$|Re(y_{12})| \ll |Im(y_{12})| \ll |Re(y_{21})|, \quad (\text{A-13})$$

$$|Im(y_{21})| \ll |Re(y_{21})|, \quad (\text{A-14})$$

$$|Re(y_{22})|, |Im(y_{22})| \ll |Re(y_{21})|. \quad (\text{A-15})$$

Verwendet man Gleichung (2-15) zur Bestimmung von f_T , dann gilt wegen (2-11) unter Berücksichtigung von (A-12) und (A-14) näherungsweise

$$f_T = f \frac{|Re(y_{21})|}{|Im(y_{11})|}. \quad (\text{A-16})$$

Mit (A-2) und (A-5) ergibt sich daraus Gleichung (2-20)

$$f_T = \frac{g_m}{2\pi(C_{GS} + C_{GD}) \left[1 + g_{DS}(R_S + R_D) + \frac{C_{GD}}{C_{GS} + C_{GD}} g_m(R_S + R_D) \right]}.$$

Für f_{\max} nach Gleichung (2-16) erhält man mit (2-12) wegen (A-13) und (A-14) näherungsweise

$$f_{\max} = f \frac{Re(y_{21})}{2[Re(y_{11})Re(y_{22}) - Re(y_{12})Re(y_{21})]^{1/2}}. \quad (\text{A-17})$$

Mit einigen grundsätzlichen Überlegungen lassen sich die Terme im Nenner von (A-17) wesentlich vereinfachen. Aus den Bedingungen (A-10), (A-11) sowie $f \ll f_T$ folgt, dass Fehler von 100% in $Re(y_{11})Re(y_{22})$ oder von 10% in $Re(y_{12})Re(y_{21})$ zu Ungenauigkeiten in f_{\max} von weniger als 10% führen. Eine solche Ungenauigkeit ist für eine Näherungsformel auf jeden Fall akzeptabel. Daher ist es sinnvoll, in (A-1) die Terme mit $(C_{GS}R_S - C_{GD}R_D)$ zu vernachlässigen, ebenso den g_{DS} -Term in (A-3) sowie den ω^2 -Term in (A-7). Damit ergibt sich aus Gleichung (A-17) unter Verwendung von (A-9) und (2-20)

$$f_{\max} = \frac{f_T}{2 \left[\frac{g_{DS}(R_G + R_S)}{N} + 2\pi f_T C_{GD} F \right]^{1/2}} \quad (\text{A-18})$$

mit

$$F = R_G \left[1 - \frac{f_T}{f_{T,i}} g_{DS}(R_S + R_D) \right] + \frac{C_{GD}}{C_{GS} + C_{GD}} R_D \frac{f_T}{f_{T,i}} N - \frac{f_T}{f_{T,i}} \frac{g_{DS}}{g_m} R_S, \quad (\text{A-19})$$

wobei

$$f_{T,i} = \frac{g_m}{2\pi(C_{GS} + C_{GD})} \quad (\text{A-20})$$

die Transitfrequenz des inneren Transistors ist. Vernachlässigt man die g_{DS} -Terme in Gleichung (A-19) und berücksichtigt außerdem, dass unter den gemachten Vorraussetzungen gilt $(f_T/f_{T,i})N \approx 1$, vereinfacht sich (A-18) zu Gleichung (2-21)

$$f_{\max} = \frac{f_T}{2 \left[\frac{g_{DS}(R_G + R_S)}{N} + 2\pi f_T C_{GD} \left(R_G + \frac{C_{GD}}{C_{GS} + C_{GD}} R_D \right) \right]^{1/2}}.$$

Die Genauigkeit von (2-20) und (2-21) lässt sich durch einen Vergleich mit numerischen Simulationen überprüfen. Als Teststruktur für diesen Vergleich diene ein Double-Gate MOSFET (Abb. 4.2) mit $L_G = 10\text{nm}$, $w = 4\text{nm}$ und $EOT = 1\text{nm}$. Für diese Struktur wurde mit dem Bauelementesimulator ATLAS eine Kleinsignalanalyse bei einer Frequenz von 40GHz durchgeführt. Für den Gleichstromarbeitspunkt bei maximalem f_T ($V_{DS} = 1\text{V}$, $V_{GS} = 0.7\text{V}$) wurden die Y-Parameter und die Kapazitäten C_{GS} und C_{GD} extrahiert sowie g_m , g_{DS} , R_S und R_D bestimmt.

Die simulierten Parameter g_m , g_{DS} , C_{GS} , C_{GD} , R_S und R_D dienten dann als Eingangsgrößen für (2-20) und (2-21). Die damit berechneten Werte für f_T und f_{\max}

befinden sich in guter Übereinstimmung mit den direkt aus den Y-Parametern unter Verwendung von (2-15) und (2-16) extrahierten Grenzfrequenzen.

Um zu überprüfen, ob die Einflüsse der verschiedenen parasitären Elemente mit (2-20) und (2-21) richtig beschrieben werden, wurden die verschiedenen Kapazitäten und Widerstände in einem weiten Bereich variiert. Als Referenz dienten hierbei Grenzfrequenzen, die aus den Y-Parametern des Gesamtvierpols aus innerem Transistor (Y-Parameter aus ATLAS) und externen Elementen berechnet wurden. Verschiebungen des Arbeitspunktes aufgrund von Spannungsabfällen über R_S und R_D bleiben dabei unberücksichtigt, was der Konsistenz des Vergleichs zugute kommt. Die Methode zur Ermittlung der Y-Parameter des Gesamtvierpols wird in Abschnitt 5.3.1 genauer beschrieben.

In Abbildung A.1 sind die mit beiden Methoden berechneten Transitfrequenzen als Funktionen der variierten Kapazität (untere Achse) und des variierten Widerstandes (obere Achse) dargestellt. Für die jeweils nicht variierten Kapazitäten wurden die mit ATLAS bestimmten Werte des inneren Transistors verwendet, nicht variierte Widerstände wurden auf einen Wert von $100\Omega\mu\text{m}$ gesetzt. Die Variationen von R_D und R_G sind in Abbildung A.1 nicht gezeigt, da zum einen die Wirkung von R_S und R_D auf f_T identisch ist und zum anderen f_T nicht von R_G abhängt. Für alle gezeigten Variationen stimmen die mit (2-20) berechneten f_T -Werte sehr gut mit den Referenzwerten überein. Gleichung (2-20) ist also in jedem Fall eine sehr gute Näherung für f_T .

Die berechneten maximalen Schwingfrequenzen sind in Abbildung A.2 als Funktion des variierten Widerstandes und in Abbildung A.3 als Funktion der variierten Kapazität zu sehen. Trotz der stärkeren Annahmen und Vereinfachungen bei der Herleitung liefert auch Gleichung (2-21) sehr gute Ergebnisse im Vergleich zu den Referenzwerten. Erstaunlicherweise stellt (2-21) auch für solche Fälle eine gute Näherung dar, in denen die Voraussetzungen für die Gültigkeit zum Teil nicht mehr erfüllt sind. Beispielsweise ist bei Widerständen von $500\Omega\mu\text{m}$ die Forderung (A-10) verletzt.

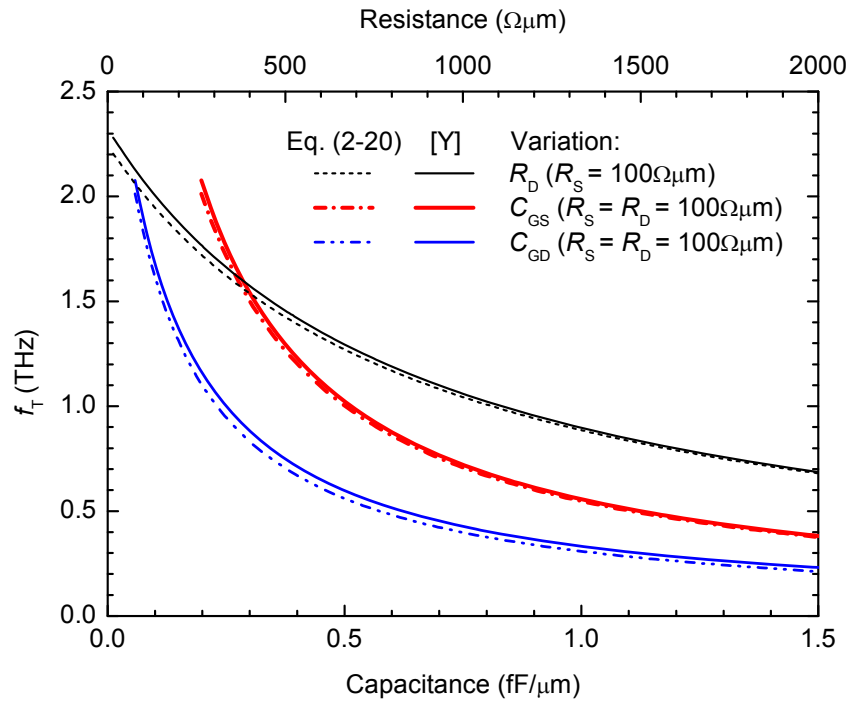


Abbildung A.1 Mit Gleichung (2-20) berechnete f_T eines DG MOSFET als Funktion von C_{GS} , C_{GD} (untere x -Achse) und R_D (obere x -Achse) im Vergleich zu direkt aus den Y-Parametern extrahierten f_T .

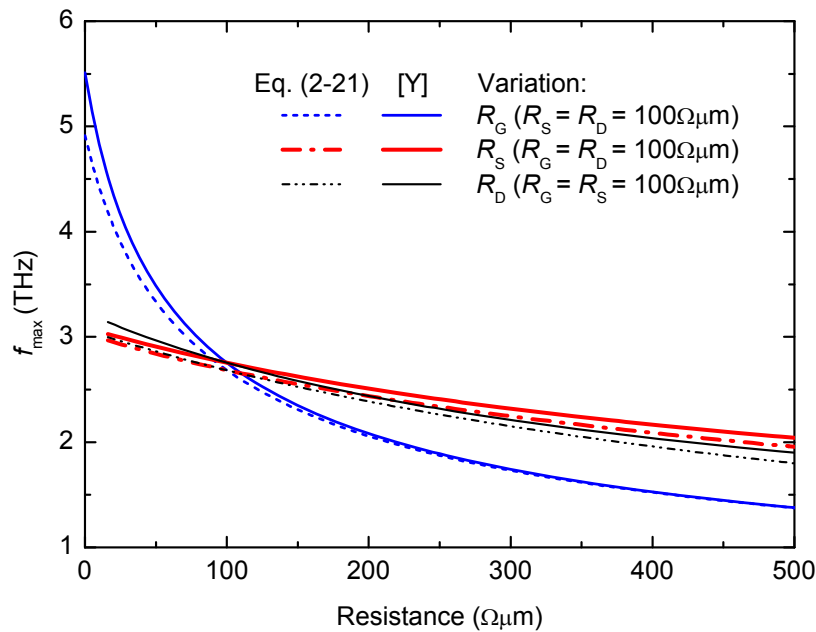


Abbildung A.2 Mit Gleichung (2-21) berechnete f_{\max} des DG MOSFET als Funktion von R_G , R_S und R_D im Vergleich zu direkt aus den Y-Parametern extrahierten f_{\max} .

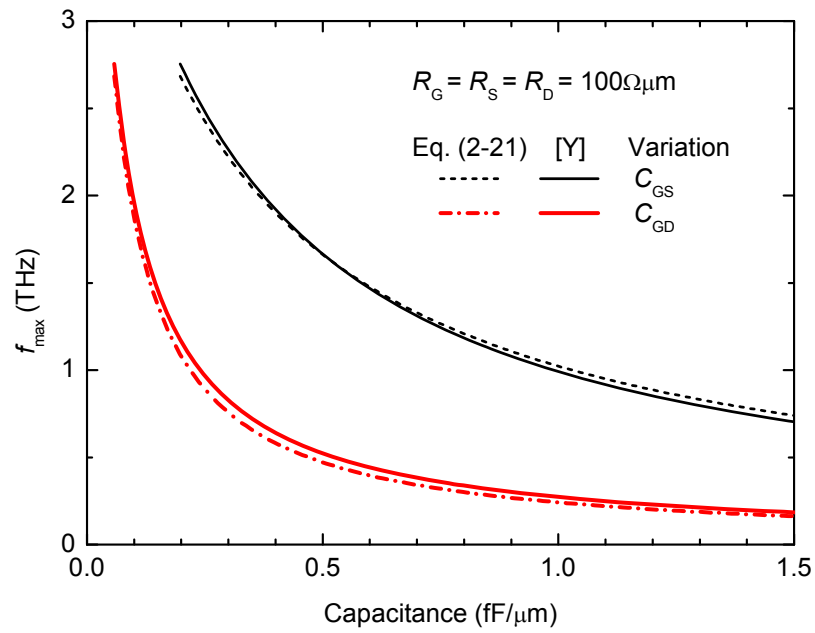


Abbildung A.3 Mit Gleichung (2-21) berechnete f_{\max} des DG MOSFET als Funktion von C_{GS} und C_{GD} im Vergleich zu direkt aus den Y-Parametern extrahierten f_{\max} .

B. Partielle Ableitungen für Gleichung (3-36)

Die partiellen Ableitungen von $\Delta\phi^{\text{QM}}$ nach den Dimensionen des Kanalquerschnitts lassen sich für MG MOSFETs aus Gleichung (3-31) und für SG bzw. DG MOSFETs aus Gleichung (3-35) analytisch bestimmen. Sie haben die Form

$$\frac{\partial \Delta\phi^{\text{QM}}}{\partial s} = \frac{k_B T}{q} \left(\frac{1}{s} - \frac{1}{\alpha} \frac{\partial \alpha}{\partial s} \right), \quad (\text{B-1})$$

wobei s für eine der Abmessungen w_{Si} bzw. t_{Si} steht. Die Ableitungen der Funktion α , Gleichung (3-32) des MG-Modells, nach w_{Si} bzw. t_{Si} lauten

$$\frac{\partial \alpha}{\partial w_{\text{Si}}} = \frac{h^2}{4k_B T w_{\text{Si}}^3} \sum_k \sum_m \sum_n \frac{N_k^{1\text{D}} m^2}{m_{xk}^*} \exp\left(-\frac{E_{k,mn}}{k_B T}\right) \quad (\text{B-2})$$

sowie

$$\frac{\partial \alpha}{\partial t_{\text{Si}}} = \frac{h^2}{4k_B T t_{\text{Si}}^3} \sum_k \sum_m \sum_n \frac{N_k^{1\text{D}} n^2}{m_{yk}^*} \exp\left(-\frac{E_{k,mn}}{k_B T}\right). \quad (\text{B-3})$$

Im Fall des SG/DG-Modells gilt $t_{\text{Si}} \rightarrow \infty$ und in (B-1) muss α durch die Funktion β [Gleichung (3-35)] ersetzt werden. Die Ableitungen von β sind

$$\frac{\partial \beta}{\partial t_{\text{Si}}} = 0 \quad \text{sowie} \quad \frac{\partial \beta}{\partial w_{\text{Si}}} = \frac{h^2}{4k_B T w_{\text{Si}}^3} \sum_k \sum_n \frac{N_k^{2\text{D}} n^2}{m_{xk}^*} \exp\left(-\frac{E_{k,n}}{k_B T}\right). \quad (\text{B-4})$$

C. Quantenkapazität von Silizium-Nanodrähten mit Bulk-Eigenschaften

Ähnlich wie für ein 1D Elektronensystem lässt sich auch für einen (hypothetischen) Silizium-Nanodraht mit Bulk-Eigenschaften eine Quantenkapazität definieren. Eine solche Definition ist jedoch nicht so eindeutig möglich wie im 1D Fall. Das liegt daran, dass in Bulk-Silizium die lokale Elektronendichte vom energetischen Abstand Leitbandkante – Ferminiveau, $E_C - E_F$, abhängt. Bei starker Inversion ist jedoch eine Bandverbiegung im Silizium unvermeidlich und der Abstand $E_C - E_F$ ändert sich innerhalb des Siliziumquerschnitts. Man muss deshalb einen mittleren oder effektiven Abstand $E_C - E_F$ annehmen, welcher nicht eindeutig definiert ist. Im 1D Elektronensystem, zum Vergleich, sind die Subbandkanten flach und der energetische Abstand zum Fermi-niveau ist konstant im gesamten Querschnitt.

Für einen Vergleich von klassischen und quantenmechanischen Quantenkapazitäten ist es daher zweckmäßig, das Problem zu vereinfachen und flache Bänder anzunehmen. In diesem Fall lässt sich die Quantenkapazität eines 3D Elektronensystems definieren als

$$C_q^{\text{bulk}} = q \frac{\partial n_{1D}^{\text{bulk}}}{\partial \varphi}, \quad (\text{C-1})$$

wobei n_{1D}^{bulk} die Elektronenlinienkonzentration ist und φ das Potential, definiert als $\varphi = -(E_C - E_F)/q$. Die Elektronenlinienkonzentration im Bulk-Siliziumkanal ist gegeben durch

$$n_{1D}^{\text{bulk}} = t_{\text{Si}} w_{\text{Si}} \int_0^{\infty} g_{3D}(\varepsilon) f(\varepsilon - q\varphi) d\varepsilon, \quad (\text{C-2})$$

wobei ε die Energie relativ zur Leitbandkante ist,

$$g_{3D}(\varepsilon) = \frac{vm_d^* \sqrt{2m_d^* \varepsilon}}{\pi^2 \hbar^3} \quad (\text{C-3})$$

ist die 3D Zustandsdichte (DOS) und

$$f(\varepsilon - q\varphi) = \frac{1}{1 + \exp\left(\frac{\varepsilon - q\varphi}{k_B T}\right)} \quad (\text{C-4})$$

die Fermi-Dirac-Verteilung. Entsprechend (C-1) erhält man durch Differentiation von Gleichung (C-2) für die Quantenkapazität eines Bulk-Silizium-Nanodrahts

$$C_q^{\text{bulk}} = \frac{\nu q^2 t_{\text{Si}} w_{\text{Si}} m_d^* \sqrt{2m_d^*}}{4k_{\text{B}} T \pi^2 \hbar^3} \int_0^\infty \sqrt{\varepsilon} \cdot \cosh^{-2} \left(\frac{\varepsilon - q\varphi}{k_{\text{B}} T} \right) d\varepsilon, \quad (\text{C-5})$$

was äquivalent zu Gleichung (3-46) ist.

Literaturverzeichnis

- [1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics* 38, no. 8, pp. 114–117, 1965.
- [2] F. Schwierz, H. Wong, and J. J. Liou, "Nanometer CMOS", Pan Stanford Publishing, 2010.
- [3] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *IEEE Journal of Solid-State Circuits* SC-9, no. 5, pp. 256-68, 1974.
- [4] The International Technology Roadmap for Semiconductors – <http://public.itrs.net/>, 2010.
- [5] ATLAS User's Manual – Device Simulation Software, Silvaco International, 2010.
- [6] S. E. Laux and M. V. Fischetti, "DAMOCLES User's Guide," IBM T. J. Watson Research Center, 1999, 2004.
- [7] White paper: Introduction to Intel's 32nm Process Technology, www.intel.com/technology/architecture-silicon/32nm/, 2009.
- [8] AMD Phenom™ II Processor Model Number and Feature Comparisons, www.amd.com, 2010.
- [9] Jing Wang, Eric Polizzi, and Mark Lundstrom, "A three-dimensional quantum simulation of silicon nanowire transistors with the effective-mass approximation," *J. Appl. Phys.* 96, pp. 2192-2203, 2004.
- [10] N. Singh, K. D. Buddharaju, S. K. Manhas, A. Agarwal, S. C. Rustagi, G. Q. Lo, N. Balasubramanian, and D.-L. Kwong, "Si, SiGe Nanowire Devices by Top-Down Technology and Their Applications," *IEEE Trans. Electron Devices* 55, no. 11, pp. 3107–3118, 2008.
- [11] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, and J. Bokor, "FinFET – A Self-Aligned Double-Gate MOSFET Scaleable to 20nm," *IEEE Trans. Electron Devices* 47, no. 12, pp. 2320-2325, 2000.
- [12] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "Sub-50 nm P-Channel FinFET," *IEEE Trans. Electron Devices* 48, no. 5, pp. 880-886, 2001.
- [13] Y. Liu, T. Matsukawa, K. Endo, M. Masahara, K. Ishii, S. O'uchi, H. Yamauchi, J. Tsukada, Y. Ishikawa, and E. Suzuki, "Advanced FinFET CMOS Technology: TiN-Gate, Fin-Height Control and Asymmetric Gate Insulator Thickness 4T-FinFETs," *IEDM Tech. Dig.*, pp. 1-4, 2006.
- [14] M. M. Hussain, C. E. Smith, H. R. Harris, C. D. Young, H.-H. Tseng, and R. Jammy, "Gate-First Integration of Tunable Work Function Metal Gates of Different Thicknesses Into High- k /Metal Gates CMOS FinFETs for Multi- V_{Th} Engineering," *IEEE Trans. Electron Devices* 57, no. 3, pp. 626-631, 2010.
- [15] K. Endo, S. O'uchi, Y. Ishikawa, Y. Liu, T. Matsukawa, K. Sakamoto, J. Tsukada, H. Yamauchi, and M. Masahara, "High Performance Fully-Depleted Tri-Gate CMOS Transistors," *IEEE Electron Device Lett.* 31, no. 6, pp. 546–548, 2010.

-
- [16] B. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High Performance Fully-Depleted Tri-Gate CMOS Transistors," *IEEE Electron Device Lett.* 24, no. 4, pp. 263-265, 2003.
 - [17] M. Lemme, T. Mollenhauer, W. Henschel, T. Wahlbrink, M. Heuser, M. Baus, O. Winkler, B. Spangenberg, R. Granzner, F. Schwier, and H. Kurz, "Influence of Channel Width on n- and p-type Nano-Wire-MOSFETs on Silicon on Insulator Substrate," *Microelectron. Eng.* 67-68, pp. 810-817, 2003.
 - [18] J.-H. Yang, Y.-S. Jin, H.-R. Lee, K.-S. Rha, J.-A. Choi, S.-K. Bae, S. Maeda, Y.-W. Kim, and K.-P. Suh, "Fully Working 1.25 μm^2 6T-SRAM Cell with 45nm Gate Length Triple Gate Transistors," *IEDM Tech. Dig.*, pp. 23-26, 2003.
 - [19] V. Gudmundsson, P.-E. Hellström, J. Luo, J. Lu, S.-L. Zhang, and M. Östling, "Fully Depleted UTB and Trigate N-Channel MOSFETs Featuring Low-Temperature PtSi Schottky-Barrier Contacts With Dopant Segregation," *IEEE Electron Device Lett.* 50, no. 5, pp. 541-543, 2009.
 - [20] M. Guillorn, J. Chang, A. Pyzyna, S. Engelmann, E. Joseph, B. Fletcher, C. Cabral, Jr., C.-H. Lin, A. Bryant, M. Darnon, J. Ott, C. Lavoie, M. Frank, L. Gignac, J. Newbury, C. Wang, D. Klaus, E. Kratschmer, J. Bucchignano, B. To, W. Graham, I. Lauer, E. Sikorski, S. Carter, V. Narayanan, N. Fuller, Y. Zhang and W. Haensch, "Trigate 6T SRAM Scaling to 0.06 μm^2 ," *IEDM Tech. Dig.*, pp. 961-963, 2009.
 - [21] M. Saitoh, Y. Nakabayashi, H. Itokawa, M. Murano, I. Mizushima, K. Uchida, and T. Numata, "Short-Channel Performance and Mobility Analysis of <110>- and <100>-Oriented Tri-Gate Nanowire MOSFETs with Raised Source/Drain Extensions," *Symp. VLSI Technology Dig.*, pp. 169-170, 2010.
 - [22] N. Singh, A. Agarwal, L. K. Bera, T. Y. Liow, R. Yang, S. C. Rustagi, C. H. Tung, R. Kumar, G. Q. Lo, N. Balasubramanian, and D.-L. Kwong, "High-Performance Fully Depleted Silicon Nanowire (Diameter ≤ 5 nm) Gate-All-Around CMOS Devices," *IEEE Electron Device Lett.* 27, no. 5, pp. 383-386, 2006.
 - [23] K. H. Yeo, S. D. Suk, M. Li, Y.-Y. Yeoh, K. H. Cho, K.-H. Hong, S. K. Yun, M. S. Lee, N. Cho, K. Lee, D. Hwang, B. Park, D.-W. Kim, D. Park, and B.-I. Ryu, "Gate-All-Around (GAA) Twin Silicon Nanowire MOSFET (TSNWFET) with 15 nm Length Gate and 4 nm Radius Nanowires," *IEDM Tech. Dig.*, pp. 1 - 4, 2006.
 - [24] B. Yang, K. D. Buddharaju, S. H. G. Teo, N. Singh, G. Q. Lo, and D. L. Kwong, "Vertical Silicon-Nanowire Formation and Gate-All-Around MOSFET," *IEEE Electron Device Lett.* 29, no. 7, pp. 791 - 794, 2008.
 - [25] J.-W. Han, D.-I. Moon, and Y.-K. Choi, "High Aspect Ratio Silicon Nanowire for Stiction Immune Gate-All-Around MOSFETs," *IEEE Electron Device Lett.* 30, no. 8, pp. 864-866, 2009.
 - [26] S. Bangsaruntip, G. M. Cohen, A. Majumdar, Y. Zhang, S. U. Engelmann, N. C. M. Fuller, L. M. Gignac, S. Mittal, J. S. Newbury, M. Guillorn, T. Barwicz, L. Sekaric, M. M. Frank, and J. W. Sleight, "High Performance and Highly Uniform Gate-All-Around Silicon Nanowire MOSFETs with Wire Size Dependent Scaling," *IEDM Tech. Dig.*, pp. 297-300, 2009.
 - [27] J. Chen, T. Saraya, and T. Hiramoto, "Mobility Enhancement over Universal Mobility in (100) Silicon Nanowire Gate-All-Around MOSFETs with Width and Height of Less Than 10nm Range," *Symp. VLSI Technology Dig.*, pp. 175-176, 2010.

- [28] Intel 22nm technology: Introducing the world's first 3-d transistor ready for high-volume manufacturing, www.intel.com/content/www/us/en/silicon-innovations/intel-22nm-technology.html, 2011.
- [29] Y. Taur and T. H. Ning, "Fundamentals of Modern VLSI Devices," Cambridge University Press, Cambridge 1998.
- [30] M. Bohr, "MOS Transistors: Scaling and Performance Trends," Semiconductor International, pp. 75-79, June 1995.
- [31] S. J. Mason, "Power Gain in Feedback Amplifier," IRE Trans. Circuit Theory, CT-1, pp. 20-25, 1954.
- [32] M. Je, I. Kwon, H. Shin, and K. Lee, "MOSFET Modeling and Parameter Extraction for RF IC's," International Journal of High Speed Electronics and Systems, vol. 11, no. 4, pp. 953-1006, 2001.
- [33] D. Lovelace, J. Costa and N. Camilleri, "Extracting Small-Signal Model Parameters of Silicon MOSFET Transistors," IEEE MTT-S Digest, pp. 865-868, 1994.
- [34] L. E. Larson, Silicon Technology Tradeoffs for Radio-Frequency/Mixed-Signal "Systems-on-a-Chip", IEEE Trans. Electron Devices 50, no. 3, pp. 683-699, 2003.
- [35] F. Schwierz and J. J. Liou, "Modern Microwave Transistors", J. Wiley & Sons, 2003.
- [36] S. M. Sze, "Physics of Semiconductor Devices", 2nd Edition, John Wiley & Sons, New York, 1981.
- [37] P. J. Tasker and B. Hughes, "Importance of source and drain resistance to the maximum f_T of millimeter-wave MODFETs," IEEE Electron Device Lett. 10, no. 7, pp. 291-293, 1989.
- [38] P. Wolf, "Microwave Properties of Schottky-Barrier Field-Effect Transistors," IBM J. Res. Develop. 14, no. 2, pp. 125-141, 1970.
- [39] S. Ohkawa, K. Suyama, and H. Ishikawa, "Low Noise GaAs Field-Effect Transistors," Fujitsu Sci. Techn. J., vol. 11, pp. 151-173, 1975.
- [40] C. E. Weitzel, "Silicon Carbide High Frequency Devices," Materials Science Forum, vols. 264-268, pp. 907-912, 1998.
- [41] L. D. Yau, "A Simple Theory to Predict the Threshold Voltage of Short-Channel IGFET's," Solid-State Electron. 17, no. 10, pp. 1059-1063, 1974.
- [42] A. Möschwitzer, "Grundlagen der Halbleiter- und Mikroelektronik", Band 1, Carl Hanser Verlag, München, 1992.
- [43] R. R. Troutman, "VLSI Limitations from Drain-Induced Barrier Lowering," IEEE Trans. Electron Devices 26, no. 4, pp. 461- 469, 1979.
- [44] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized Scaling Theory and Its Application to a 1/4 micrometer MOSFET design," IEEE Trans. Electron Devices 31, no. 4, pp. 452-462, 1984.
- [45] Y. Taur, D. A. Buchanan, W. Chen, et.al, "CMOS Scaling into the Nanometer Regime," Proc. IEEE, vol. 85, no. 4, pp. 486-503, 1997.

-
- [46] H. Wong and Y. Taur, "Three-Dimensional "Atomistic" Simulation of Discrete Random Dopant Distribution Effects in Sub-0.1 μ m MOSFET's," IEDM Tech. Dig., pp. 705 - 708, 1993.
 - [47] A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, and S. Saini, "Quantum Mechanical Enhancement of the Random Dopant Induced Threshold Voltage Fluctuations and Lowering in Sub 0.1 Micron MOSFETs," IEDM Tech. Dig., pp. 535 - 538, 1999.
 - [48] A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, and S. Saini, "Increase in the Random Dopant Induced Threshold Voltage Fluctuations and Lowering in Sub-100nm MOSFETs Due to Quntum Effects: A 3-D Density-Gradient Simulation Study," IEEE Trans. Electron Devices 48, no. 4, pp. 722-729, 2001.
 - [49] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Boeuf, S. Monfray, F. Payet, A. Pouydebasque, M. Szczap, A. Farcy, F. Arnaud, S. Clerc, M. Sellier, A. Cathignol, J.-P. Schoellkopf, E. Perea, R. Ferrant, and H. Mingam, "Innovative Materials, Devices, and CMOS Technologies for Low-Power Mobile Multimedia," IEEE Trans. Electron Devices 55, no. 1, pp. 96-130, 2008.
 - [50] C. Y. Wong, J. Y. Sun, Y. Taut, C. S. Oh, R. Angelucci, and B. Davari, "Doping of N⁺ and P⁺ polysilicon in a dual-gate CMOS process," IEDM Tech. Dig., pp. 238 - 241, 1988.
 - [51] E. Josse and T. Skotnicki, "2D analysis of gate polydepletion in ultra short MOSFETs," Proc. ESSDERC, pp. 207-210, 2001.
 - [52] S.-H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, "Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFET's," IEEE Electron Device Letters, vol. 18, no. 5, pp. 209-211, 1997.
 - [53] J. Robertson, "High dielectric constant gate oxides for metal oxide Si transistors," Rep. Prog. Phys. 69, pp. 327-396, 2006.
 - [54] A. S. Oates, "Reliability issues for high-K gate dielectrics," IEDM Tech. Dig., pp. 923 - 926, 2003.
 - [55] D.-L. Kwong, "CMOS integration issues with high-k gate stack," Proc. 11th IPFA, pp. 17-20, 2004.
 - [56] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, and G. Ghibaudo, "Review on high-k dielectrics reliability issues," IEEE Trans. Device and Materials Reliab., vol. 5, no. 1, pp. 5 - 19, 2005.
 - [57] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering," J. Appl. Phys. 90, pp. 4587- 4608, 2001.
 - [58] R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz, "High- κ / Metal-Gate Stack and Its MOSFET Characteristics," IEEE Electron Device Letters, vol. 25, no. 6, pp. 408- 410, 2004.
 - [59] M. T. Bohr, R. Chau, T. Ghani, and K. Mistry, "The High- k Solution," IEEE Spectrum, pp. 29- 35, Oct. 2007.
 - [60] J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?," IEDM Tech. Dig., pp. 707 - 710, 2002.
 - [61] M. Städele, "Influence of Source-Drain Tunneling on the Subthreshold Behavior of sub-10nm Double-gate MOSFETs," Proc. ESSDERC, pp. 135 - 138, 2002.

- [62] M. Bescond, J.L. Autran, D. Munteanu, N. Cavassilas, and M. Lannoo, "Atomic-scale Modeling of Source-to-Drain Tunneling in Ultimate Schottky Barrier Double-Gate MOSFETs," *Proc. ESSDERC*, pp. 395 - 398, 2003.
- [63] M. Bescond, K. Nehari, J. L. Autran, N. Cavassilas, D. Munteanu, and M. Lannoo, "3D Quantum Modelling and Simulation of Multiple-Gate Nanowire MOSFETs," *IEDM Tech. Dig.*, pp. 617-620, 2004.
- [64] H.-S. P. Wong, D. J. Frank, and P. M. Solomon, "Device Design Considerations for Double-Gate, Ground-Plane, and Single-Gated Ultra-Thin SOI MOSFET's at the 25nm Channel Length Generation," *IEDM Tech. Dig.*, pp. 407-410, 1998.
- [65] V. P. Trivedi and J. G. Fossum, "Quantum-Mechanical Effects on the Threshold Voltage of Undoped Double-Gate MOSFETs," *IEEE Electron Device Lett.* 26, no. 8, pp. 579-582, 2005.
- [66] R. Granzner, F. Schwierz, and V. M. Polyakov, "An Analytical Model for the Threshold Voltage Shift Caused by Two-Dimensional Quantum Confinement in Undoped Multiple-Gate MOSFETs," *IEEE Trans. Electron Devices* 54, no. 9, pp. 2562-2565, 2007.
- [67] R. Herrmann und U. Preppernau, "Elektronen im Kristall", Akademie-Verlag, Berlin, 1979.
- [68] S. Selberherr, "Analysis and Simulation of Semiconductor Devices", Springer-Verlag, Wien, 1984.
- [69] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, "A Review of Hydrodynamic and Energy-Transport Models for Semiconductor Device Simulation," *Proc. IEEE*, vol. 91, no. 2, pp. 251- 274, 2003.
- [70] D. Nuernbergk, "Simulation des Transportverhaltens in Si/Si_{1-x}Ge_x/Si-Heterobipolartransistoren", Dissertation, TU Ilmenau, 1997.
- [71] R. Stratton, "Diffusion of hot and cold electrons in semiconductor barriers," *Phys. Rev.*, vol. 126, no. 6, pp. 2002-2014, 1962.
- [72] R. Stratton, "Semiconductor Current-Flow Equations (Diffusion and Degeneracy)," *IEEE Trans. Electron Devices* 19, no. 12, pp. 1288-1292, 1972.
- [73] J.-H. Rhew, Z. Ren, and M. Lundstrom, "Numerical study of a ballistic MOSFET," *Solid-State Electron.* 46, pp. 1899-1906, 2002.
- [74] H. Kosina, M. Nedjalkov, and S. Selberherr, "Theory of the Monte Carlo Method for Semiconductor Device Simulation," *IEEE Trans. Electron Devices* 47, no. 10, pp. 1898-1908, 2000.
- [75] F. M. Bufler, Y. Asahi, H. Yoshimura, C. Zechner, A. Schenk, and W. Fichtner, "Monte Carlo Simulation and Measurement of Nanoscale n-MOSFETs," *IEEE Trans. Electron Devices* 50, no. 2, pp. 418-424, 2003.
- [76] S. Datta, "Nanoscale device modeling: the Green's function method," *Superlattices and Microstructures*, vol. 28, no. 4, pp. 253-278, 2000.
- [77] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, "Two-dimensional quantum mechanical modeling of nanotransistors," *J. Appl. Phys.* 91, pp. 2343-2354, 2002.

-
- [78] R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom, and D. Jovanovic, "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches," *J. Appl. Phys.* 92, pp. 3730–3739, 2002.
 - [79] Z. Ren, R. Venugopal, S. Goasguen, S. Datta, and M. S. Lundstrom, "nanoMOS 2.5: A Two-Dimensional Simulator for Quantum Transport in Double-Gate MOSFETs," *IEEE Trans. Electron Devices* 50, no. 9, pp. 1914–1925, 2003.
 - [80] A. Martinez, J. R. Barker, A. Svizhenko, A. Anantram, M. Bescond, and A. Asenov, "Development of a Full 3D NEGF Nano-CMOS Simulator," *Proc. SISPAD 2006*, pp. 253–256, 2006.
 - [81] R. Lake and S. Datta, "Nonequilibrium Green's function method applied to double barrier resonant-tunneling diodes," *Phys. Rev B*, vol. 45, p.6670, 1992.
 - [82] M. J. Gilbert, R. Akis, and D. K. Ferry, "Phonon-assisted ballistic to diffusive crossover in silicon nanowire transistors," *J. Appl. Phys.* 98, 094303, 2005.
 - [83] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, and M. Lundstrom, "A simple quantum mechanical treatment of scattering in nanoscale transistors," *J. Appl. Phys.*, vol. 93, p. 5613, 2003.
 - [84] S. Datta, "Quantum transport: atom to transistor", Cambridge Univ. Press, 2006.
 - [85] A. Svizhenko and M. P. Anantram, "Role of Scattering in Nanotransistors," *IEEE Trans. Electron Devices* 50, no. 6, pp. 1459–1466, 2003.
 - [86] M. Büttiker, "Four-terminal phase coherent conductance," *Phys. Rev. Lett.*, vol. 57, p. 1761, 1986.
 - [87] V. Sverdlov, H. Kosina, S. Selberherr, "Modeling current transport in ultra-scaled field-effect transistors," *Microelectronics Reliab.*, vol. 47, pp. 11–19, 2007.
 - [88] V. Sverdlov, A. Gehring, H. Kosina, S. Selberherr, "Quantum transport in ultra-scaled double-gate MOSFETs: A Wigner function-based Monte Carlo approach," *Solid-State Electron.* 49, pp. 1510–1515, 2005.
 - [89] M. J. van Dort, P. H. Woerlee, and A. J. Walker, "A simple model for quantization effects in heavily-doped silicon MOSFET's at inversion conditions," *Solid-State Electron.* 37, no. 3, pp. 411–414, 1994.
 - [90] A. Wettstein, A. Schenk, and W. Fichtner, "Quantum Device-Simulation with the Density-Gradient Model on Unstructured Grids," *IEEE Trans. Electron Devices* 48, no. 2, pp. 279–284, 2001.
 - [91] R. Granzner, V. M. Polyakov, F. Schwierz, M. Kittler, and T. Doll, "On the Suitability of DD and HD Models for the Simulation of Nanometer Double-Gate MOSFETs," *Physica E* 19, pp. 33–38, 2003.
 - [92] R. Granzner, V. M. Polyakov, F. Schwierz, M. Kittler, R. J. Luyken, W. Rösner, and M. Städele, "Simulation of Nanoscale MOSFETs Using Modified Drift-Diffusion and Hydrodynamic Models and Comparison with Monte Carlo Results," *Microelectron. Eng.* 83, pp. 241–246, 2006.
 - [93] S. E. Laux, M. V. Fischetti, and D. J. Frank, "Monte Carlo Analysis of Semiconductor Devices: The DAMOCLES Program," *IBM J. Res. Develop.* 34, pp. 466–494, 1990.

- [94] C. Lombardi, S. Manzini, A. Saporito, and M. Vanzi, "A Physically Based Mobility Model for Numerical Simulation of Nonplanar Devices," *IEEE Trans. Computer Aided Design* 7, pp. 1164-1171, 1988.
- [95] D. M. Caughey and R. E. Thomas, "Carrier Mobilities in Silicon Empirically Related to Doping and Field," *Proc. IEEE* 52, pp. 2192-2193, 1967.
- [96] K. Banoo and M. S. Lundstrom, "Electron Transport in a Model Si Transistor," *Solid-State Electron.* 44, pp. 1689-1695, 2000.
- [97] J.-H. Rhew, Z. Ren, and M. Lundstrom, "Benchmarking Macroscopic Transport models for nanotransistor TCAD," *J. Comp. Electronics* 1, pp. 385-388, 2002.
- [98] Z. Ren and M. Lundstrom, "Simulation of nanoscale MOSFETs: a scattering theory interpretation," *Superlattices and Microstructures*, vol. 27, no. 2/3, pp. 177-189, 2000.
- [99] M. Lundstrom, "Elementary Scattering Theory of the Si MOSFET," *IEEE Electron Device Lett.* 18, no. 7, pp. 361-363, 1997.
- [100] M. S. Lundstrom, "On the Mobility Versus Drain Current Relation for a Nanoscale MOSFET," *IEEE Electron Device Lett.* 22, no. 6, pp. 293-295, 2001.
- [101] M. Lundstrom and Z. Ren, "Essential Physics of Carrier Transport in Nanoscale MOSFETs," *IEEE Trans. Electron Devices*. 49, no. 1, pp. 133-141, 2002.
- [102] A. Rahman and M. S. Lundstrom, "A Compact Scattering Model for the Nanoscale Double-Gate MOSFET," *IEEE Trans. Electron Devices* 49, no. 3, pp. 481-489, 2002.
- [103] F. Assad, Z. Ren, S. Datta, and M. Lundstrom, "Performance Limits of Silicon MOSFETs," *IEDM Tech. Dig.*, pp. 547-550, 1999.
- [104] J. D. Bude, "MOSFET Modeling Into the Ballistic Regime," *Proc. SISPAD*, pp. 23-26, 2000.
- [105] H. M. Nayfeh, J. L. Hoyt, and D. A. Antoniadis, "Investigation of Scaling Methodology for Strained Si n-MOSFETs Using a Calibrated Transport Model," *IEDM Tech. Dig.*, pp. 475-478, 2003.
- [106] S. E. Laux, "Techniques for Small-Signal Analysis of Semiconductor Devices," *IEEE Trans. CAD*, vol. CAD-4, no. 4, pp. 472-481, 1985.
- [107] G. W. Taylor, W. Fichtner, and J. G. Simmons, "A Discription of MOS Internodal Capacitances for Transient Simulations," *IEEE Trans. CAD*, vol CAD-1, no. 4, pp. 150-156, 1982.
- [108] J. J. H. van den Biesen, "A Simple Regional Analysis of Transit Times in Bipolar Transistors," *Solid-State Electron.*, vol. 29, no. 5, pp. 529-534, 1986.
- [109] B. Heinemann, "2D-Bauelementesimulation der elektrischen Eigenschaften von SiGe-HBTs", *Dissertation*, TU Berlin, 1997.
- [110] S. Zhou, J. Jiang, and Q. Cai, "Numerical analysis of the cut-off frequency of ultra-small ballistic double-gate MOSFETs 2D nonequilibrium Green's function approach," *Solid-State Electron.* 49, pp. 1951-1955, 2005.

-
- [111] R. Wang, J. Zhuge, R. Huang, Y. Tian, H. Xiao, L. Zhang, C. Li, X. Zhang, and Y. Wang, "Analog/RF Performance of Si Nanowire MOSFETs and the Impact of Process Variation," *IEEE Trans. Electron Devices* 54, no. 6, pp. 1288-1294, 2007.
 - [112] G. Baccarani and S. Reggiani, "A Compact Double-Gate MOSFET Model Comprising Quantum-Mechanical and Nonstatic Effects," *IEEE Trans. Electron Devices* 46, no. 8, pp. 1656-1666, 1999.
 - [113] J. L. Autran, K. Nehari, and D. Munteanu, "Compact Modelling of the Threshold Voltage in Silicon Nanowire MOSFET Including 2D-Quantum Confinement Effects," *Molecular Simulation*, vol. 31, no. 12, pp. 839-843, 2005.
 - [114] M. Bescond, N. Cavassilas, K. Kalna, K. Nehari, L. Raymond, J. L. Autran, M. Lannoo, and A. Asenov, "Ballistic transport in Si, Ge, and GaAs nanowire MOSFETs," *IEDM Tech. Dig.*, pp. 533-536, 2005.
 - [115] F. Stern and W. E. Howard, "Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit," *Phys. Rev.*, vol. 163, no. 3, pp. 816-835, 1967.
 - [116] M. Lundstrom and J. Guo, "Nanoscale Transistors: Device Physics, Modeling and Simulation," Springer Science + Business Media, Inc., New York, 2006.
 - [117] SCHRED ist ein eindimensionaler numerischer Simulator zur selbstkonsistenten Lösung von Schrödinger- und Poissongleichung. Online erreichbar unter nanohub.org/tools/schred.
 - [118] J.-P. Colinge, J. C. Alderman, W. Xiong, and C. R. Cleavelin, "Quantum-Mechanical Effects in Trigate SOI MOSFETs," *IEEE Trans. Electron Devices* 53, no. 5, pp. 1131-1136, 2006.
 - [119] Q. Chen, E. M. Harrell, II, and J. D. Meindl, "A Physical Short-Channel Threshold Voltage Model for Undoped Symmetric Double-Gate MOSFETs," *IEEE Trans. Electron Devices* 50, no. 7, pp. 1631-1637, 2003.
 - [120] S. Takagi and A. Toriumi, "Quantitative Understanding of Inversion-Layer Capacitance in Si MOSFETs," *IEEE Trans. Electron Devices*, vol. 42, no. 12, pp. 2125 - 2130, 1995.
 - [121] S. Luryi, "Quantum capacitance devices," *Appl. Phys. Lett.* 52, 501, 1988.
 - [122] D. Vasileska, D. K. Schroder, and D. K. Ferry, "Scaled Silicon MOSFET's: Degradation of the Total Gate Capacitance," *IEEE Trans. Electron Devices*, vol. 44, no. 4, pp. 584 - 587, 1997.
 - [123] D. Vasileska and D. K. Ferry, "The influence of space quantization effects on the threshold voltage, inversion layer and total gate capacitances in scaled Si-MOSFETs," *Nanotechnology* 10, 192 - 197, 1999.
 - [124] H. S. Pal, K. D. Cantley, S. S. Ahmed, and M. S. Lundstrom, "Influence of Bandstructure and Channel Structure on the Inversion Layer Capacitance of Silicon and GaAs MOSFETs," *IEEE Trans. Electron Devices*, vol. 55, no. 3, pp. 904 - 908, 2008.
 - [125] D. L. John, L. C. Castro, and D. L. Pulfrey, "Quantum capacitance in nanoscale device modeling," *J. Appl. Phys.* 96, 5180, 2004.
 - [126] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Trans. Electron Devices*, vol. 50, pp. 1853-1864, Sep. 2003.
 - [127] J. Knoch, W. Riess, and J. Appenzeller, "Outperforming the Conventional Scaling Rules in the Quantum-Capacitance Limit," *IEEE Electron Device Lett.*, vol. 29, pp. 372-374, Apr. 2008.

- [128] B. Yu, L. Wang, Y. Yuan, P. M. Asbeck, and Yuan Taur, "Scaling of Nanowire Transistors," *IEEE Trans. Electron Devices*, vol. 55, pp. 2846 - 2858, Nov. 2008.
- [129] L. Wang, D. Wang, and P. M. Asbeck, "A numerical Schrödinger–Poisson solver for radially symmetric nanowire core–shell structures," *Solid State Electron.*, vol. 50, pp. 1732–1739, Nov./Dec. 2006.
- [130] F. J. G. Ruiz, A. Godoy, F. Gámiz, C. Sampedro, and L. Donetti, "A comprehensive study of the corner effects in pi-gate SOI MOSFETs including quantum effects," *IEEE Trans. Electron Devices*, vol. 54, pp. 3369–3377, Dec. 2007.
- [131] F. J. G. Ruiz, I. M. Tienda-Luna, A. Godoy, L. Donetti, and F. Gámiz, "Equivalent Oxide Thickness of Trigate SOI MOSFETs With High- κ Insulators," *IEEE Trans. Electron Devices*, vol. 56, pp. 2711 - 2719, Nov. 2009.
- [132] R. Granzner, S. Thiele, C. Schippel, and F. Schwier, "Quantum effects on the gate capacitance of trigate SOI MOSFETs", *IEEE Trans. Electron Dev.* 57, pp. 3231-3237, 2010.
- [133] nextnano³ – next generation 3D nanodevice simulator, www.nextnano.de/nextnano3.
- [134] S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, P. Vogl, "nextnano: General Purpose 3-D Simulations," *IEEE Trans. Electron Devices*, vol. 54, pp. 2137- 2142, Sep. 2007.
- [135] W. Quan, D. M. Kim, and H.-D. Lee, "Quantum C–V Modeling in Depletion and Inversion: Accurate Extraction of Electrical Thickness of Gate Oxide in Deep Submicron MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 5, pp. 889 - 894, 2002.
- [136] R. Kotlyar, B. Obradovic, P. Matagne, M. Stettler, and M. D. Giles, "Assessment of room-temperature phonon-limited mobility in gated silicon nanowires," *Appl. Phys. Lett.* 84, 5270, 2004.
- [137] S. Jin, M. V. Fischetti, and T. Tang, "Modeling of electron mobility in gated silicon nanowires at room temperature: Surface roughness scattering, dielectric screening, and band nonparabolicity," *J. Appl. Phys.* 102, 083715, 2007.
- [138] E. B. Ramayya, D. Vasileska, S. M. Goodnick, and I. Knezevic, "Electron transport in silicon nanowires: The role of acoustic phonon confinement and surface roughness scattering," *J. Appl. Phys.* 104, 063711, 2008.
- [139] Y. Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, "High Performance Silicon Nanowire Field Effect Transistors," *Nano Lett.* 3, pp. 149-152, 2003.
- [140] S.-M. Koo, A. Fujiwara, J.-P. Han, E. M. Vogel, C. A. Richter, and J. E. Bonevich, "High Inversion Current in Silicon Nanowire Field Effect Transistors," *Nano Lett.* 4, pp. 2197-2201, 2004.
- [141] A. Godoy, F. Ruiz, C. Sampedro, F. Gámiz, and U. Ravaioli, "Calculation of the phonon-limited mobility in silicon Gate All-Around MOSFETs," *Solid State Electron.* 51, pp. 1211–1215, 2007.
- [142] R. Mickevičius and V. Mitin, "Acoustic-phonon scattering in a rectangular quantum wire," *Phys. Rev. B* 48, pp. 17194–17201, 1993.
- [143] E. Fernández-Díaz, A. Alarcón, and X. Oriols, "Modeling Quantum Transport Under AC Conditions: Application to Intrinsic High-Frequency Limits for Nanoscale Double-Gate Si MOSFETs," *IEEE Trans. Nanotechnology*, vol. 4, no. 5, pp. 563-569, 2005.

-
- [144] D. Buss, B. L. Evans, J. Bellay, W. Krenik, B. Haroun, D. Leipold, K. Maggio, J.-Y. Yang, and T. Moise, "SOC CMOS Technology for Personal Internet Products," *IEEE Trans. Electron Devices* 50, no. 3, pp. 546-556, 2003.
 - [145] H. Stöcker, "Taschenbuch der Physik", 4. Auflage, Verlag Harri Deutsch, 2000.
 - [146] J. G. Fossum, "Physical insights on nanoscale multi-gate CMOS design," *Solid-State Electron.* 51, pp. 188-194, 2007.
 - [147] F. Fillot, T. Morel, S. Minoret, I. Matko, S. Maitrejean, B. Guillaumot, B. Chenevier, and T. Billon, "Investigations of titanium nitride as metal gate material, elaborated by metal organic atomic layer deposition using TDMAT and NH₃," *Microelectron. Eng.* 82, pp. 248-253, 2005.
 - [148] K. Ahmed und K. Schuegraf, "Transistor Wars – Rival Architectures Face Off in a Bid to Keep Moore's Law Alive," *IEEE Spectrum* 48, no. 11, pp. 44-49, 2011.
 - [149] S.-H. Kim, J. G. Fossum, and V. P. Trivedi, "Bulk Inversion in FinFETs and Implied Insights on Effective Gate Width," *IEEE Trans. Electron Devices* 52, no. 9, pp. 1993-1997, 2005.
 - [150] J. Wang, P. M. Solomon, and M. Lundstrom, "A General Approach for the Performance Assessment of Nanoscale Silicon FETs," *IEEE Trans. Electron Devices* 51, no. 9, pp. 1366-1370, 2004.
 - [151] S. E. Thompson, R. S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M. T. Bohr, "In Search of "Forever," Continued Transistor Scaling One New Material at a Time," *IEEE Trans. Semiconduct. Manuf.* 18, no. 1, pp. 26-36, 2005.
 - [152] B. Jagannathan, D. Chidambarrao, and J. Pekarik, "300GHz Transistor Performance in Production CMOS Technologies," *Proc. 64th Device Research Conf.*, pp. 199-200, 2006.
 - [153] S. Lee, B. Jagannathan, S. Narasimha, A. Chou, N. Zamdmer, J. Johnson, R. Williams, L. Wagner, J. Kim, J.-O. Plouchart, J. Pekarik, S. Springer, and G. Freeman, "Record RF performance of 45-nm SOI CMOS Technology," *IEDM Tech. Dig.*, pp. 255-258, 2007.
 - [154] C. H. Diaz, D. D. Tang, and J. Sun, "CMOS Technology for MS/RF SoC," *IEEE Trans. Electron Devices* 50, no. 3, pp. 557-566, 2003.
 - [155] W. Weißgerber, "Elektrotechnik für Ingenieure 3 – Ausgleichsvorgänge, Fourieranalyse, Vierpoltheorie", 4. Auflage, Vieweg, Braunschweig/Wiesbaden, 1999.
 - [156] S. Bull, "Entwicklung eines C-Programms zur Berechnung der Kleinsignalparameter, Verstärkungen und Grenzfrequenzen von Mikrowellentransistoren", Studienarbeit, TU Ilmenau, 2000.

Veröffentlichungen

Veröffentlichungen zu Nanometer-MOSFETs

- [1] R. Granzner, S. Thiele, C. Schippel, and F. Schwierz, "Quantum effects on the gate capacitance of trigate SOI MOSFETs", IEEE Trans. Electron Dev. 57, pp. 3231-3237, 2010.
- [2] J. Racko, R. Granzner, F. Schwierz, J. Breza, D. Donoval, O. Kucera, and P. Pintes, "Comparison of a Standard and a Schottky Dual Gate MOSFET", J. Electrical Eng. - Elektrotechniky Casopis 59, pp. 81-85, 2008.
- [3] R. Granzner, F. Schwierz, and V. M. Polyakov, "An Analytical Model for the Threshold Voltage Shift Caused by Two-Dimensional Quantum Confinement in Undoped Multiple-Gate MOSFETs," IEEE Trans. Electron Devices 54, no. 9, pp. 2562-2565, 2007.
- [4] R. Granzner, V. M. Polyakov, F. Schwierz, M. Kittler, R. J. Luyken, W. Rösner, and M. Städele, "Simulation of Nanoscale MOSFETs Using Modified Drift-Diffusion and Hydrodynamic Models and Comparison with Monte Carlo Results", Microelectron. Eng. 83, pp. 241-246, 2006.
- [5] M. Kittler, F. Schwierz, and R. Granzner, "Benchmarking of Straddle-Gate MOSFETs", Proc. 51th IWK, pp. 255-256, 2006.
- [6] M. Kittler, F. Schwierz, R. Granzner, M. Lemme, T. Wahlbrink, and H. Kurz, "The I_{on} - I_{off} Performance of Straddle-Gate FETs for HighPerformance Logic", Proc. ULIS, pp. 149-152, 2006.
- [7] J. Racko, D. Donoval, A. Vrbicky, R. Granzner, and F. Schwierz, "Analysis of a Double Gate Tunneling Schottky-MOSFET with Ultrathin Channel by 2-D Modeling and Simulation", Proceedings of the 11th International Workshop on Applied Physics of Condensed Matter (APCOM), Mala Lucivna (Slovakia), pp. 119-121, 2005.
- [8] M. C. Lemme, T. Mollenhauer, W. Henschel, T. Wahlbrink, M. Baus, O. Winkler, R. Granzner, F. Schwierz, B. Spangenberg, and H. Kurz, "Subthreshold Behavior of Triple-Gate MOSFETs on SOI Material", Solid-State Electronics 48, pp. 529-534, 2004.
- [9] R. Granzner, V. M. Polyakov, F. Schwierz, M. Kittler, and T. Doll, "On the Suitability of DD and HD Models for the Simulation of Nanometer Double-Gate MOSFETs," Physica E 19, pp. 33-38, 2003.
- [10] M. Lemme, T. Mollenhauer, W. Henschel, T. Wahlbrink, M. Heuser, M. Baus, O. Winkler, B. Spangenberg, R. Granzner, F. Schwierz, and H. Kurz, "Influence of Channel Width on n- and p-type Nano-Wire-MOSFETs on Silicon on Insulator Substrate," Microelectron. Eng. 67-68, pp. 810-817, 2003.
- [11] M. Kittler, R. Granzner, F. Schwierz, W. Henschel, T. Wahlbrink, and H. Kurz, "Simulation and Optimization of EJ-MOSFETs", Solid-State Electronics 47, pp. 1193-1198, 2003.
- [12] R. Granzner, V. M. Polyakov, F. Schwierz, and M. Kittler, "On the Limitation of DD and HD Models for DG SOI MOSFET by Comparison to Monte Carlo Simulation", Book of Abstracts, 4th Int. Symp. Nanostructures and Mesoscopic Systems Nano MES, Tempe, USA, pp. MA7-1 - MA7-2, Tempe, USA, 2003.

- [13] M. Lemme, T. Mollenhauer, W. Henschel, T. Wahlbrink, H. Kurz, M. Heuser, M. Baus, O. Winkler, B. Spangenberg, R. Granzner, and F. Schwierz, "Influence of Channel Width on n- and p-type Nano-Wire-MOSFETs on Silicon on Insulator Substrate", Proc. 28th International Conference on Micro- and Nano-Engineering (MNE), September 16-19, Lugano, Switzerland, 2002.
- [14] M. Heuser, M. Baus, B. Hadam, O. Winkler, B. Spangenberg, R. Granzner, M. Lemme, and H. Kurz, "Fabrication of Wire-MOSFETs on Silicon-on-Insulator Substrate", Microelectronic Engineering 61-62, pp. 613-618, 2002.
- [15] R. Granzner, F. Schwierz, M. Kittler, T. Doll, M. Heuser, and M. Baus, "A Simulation Study on Three-Dimensional Device Design of Wire-MOSFETs", Proc. 3rd Europ. Workshop Ultimate Integration of Silicon (ULIS), March 7-8, Munich, Germany, pp. 95-98, 2002.
- [16] M. Kittler, R. Granzner, and F. Schwierz, "Simulation and Optimization of EJ-MOSFETs", Proc. 3rd Europ. Workshop Ultimate Integration of Silicon (ULIS), March 7-8, Munich, Germany, pp. 175-178, 2002.
- [17] M. Heuser, M. Baus, B. Hadam, O. Winkler, B. Spangenberg, R. Granzner, M. Lemme, and H. Kurz, "Fabrication of Wire-MOSFETs on Silicon-on-Insulator Substrate", Proc. 27th International Conference on Micro- and Nano-Engineering (MNE), Sept. 16-19, Grenoble, France, p. 220, 2001.

Veröffentlichungen zu weiteren Themen

- [18] N. Al Mustafa, R. Granzner, V. M. Polyakov, J. Racko, M. Mikolášek, J. Breza, and F. Schwierz, "The Coexistence of Two-Dimensional Electron and Hole Gases in GaN-based Heterostructures", J. Appl. Phys. 111, 044512 (2012).
- [19] R. Granzner, E. Tschumak, M. Kittler, K. Tonisch, W. Jatal, J. Pezoldt, D. As, and F. Schwierz, "Vertical design of cubic GaN-based high electron mobility transistors", J. Appl. Phys. 110, 114501 (2011).
- [20] J. Racko, M. Mikolasek, P. Benko, O. Gallo, L. Harmatha, R. Granzner, and F. Schwierz, "Coupled defect level recombination in the p-n junction", J. Electrical Eng., vol. 62, no. 6, pp. 355-358 (2011).
- [21] R. Granzner, N. Mustafa, J. Racko, M. Mikolášek, J. Breza, and F. Schwierz, "On The Coexistence of Two-Dimensional Electron and Hole Gases in GaN-based Heterostructures", Proc. 17th APCOM, pp. 126-129, 2011.
- [22] J. Racko, M. Mikolasek, L. Harmatha, J. Breza, B. Hudec, K. Fröhlich, J. Aarik, A. Tarre, R. Granzner, and F. Schwierz, "Analysis of leakage current mechanisms in RuO₂-TiO₂-RuO₂ MIM structures", J. Vac. Sci. Technol. B 29, 01AC08 1-8 (2011).
- [23] M. Mikolášek, J. Racko, L. Harmatha, O. Gallo, J. Režnák, F. Schwierz, and R. Granzner, "A new model of trap assisted band-to-band tunneling", Proc. 8th ASDAM, pp. 195-198 (2010).
- [24] R. Granzner, M. Kittler, F. Schwierz, and V. M. Polyakov, "Vertical design of InN field effect transistors", Proc. ESSDERC, pp. 428-431 (2010).
- [25] J. Racko, M. Mikolasek, L. Harmatha, J. Breza, B. Hudec, K. Fröhlich, J. Aarik, A. Tarre, R. Granzner, and F. Schwierz, "Analysis of leakage mechanisms in RuO₂-TiO₂-RuO₂ MIM structures", Book of Abstracts WoDiM 2010.

- [26] D. J. As, E. Tschumak, F. Niebelschütz, W. Jatal, J. Pezoldt, R. Granzner, F. Schwierz, and K. Lischka, "Cubic AlGa_N/Ga_N hetero-field effect transistors with normally-on and normally-off operation", MRS Symp. Proc. Vol. 12020, 104-108 (2010).
- [27] K. Tonisch, W. Jatal, R. Granzner, M. Kittler, U. Baumann, F. Schwierz, and J. Pezoldt, "2H-AlGa_N/Ga_N HEMT on 3C-SiC(111)/Si(111) substrates", Materials Science Forum 645-648, pp. 1219-1222 (2010).
- [28] E. Tschumak, R. Granzner, J. K. N. Lindner, F. Schwierz, K. Lischka, H. Nagasawa, M. Abe, and D. J. As, "Nonpolar cubic AlGa_N/Ga_N heterojunction field-effect transistor on Ar⁺ implanted 3C-SiC (001)", Appl. Phys. Lett. 96, p. 253501 (2010).
- [29] J. Racko, M. Mikolasek, R. Granzner, J. Breza, D. Donoval, A. Grmanova, L. Harmatha, F. Schwierz, and K. Fröhlich, "Trap-assisted tunneling current in MIM structures", Cent. Eur. J. Phys. 9, pp. 230-241, 2010.
- [30] K. Köhler, S. Müller, R. Aidam, P. Waltereit, W. Pletschen, L. Kirste, H. P. Menner, W. Bronner, A. Leuther, R. Quay, M. Mikulla, O. Ambacher, R. Granzner, F. Schwierz, C. Buchheim, and R. Goldhahn, "Influence of the surface potential on electrical properties of Al_xGa_{1-x}N/GaN heterostructures with different Al-content: Effect of growth method", J. Appl. Phys. 107, 053711 (2010).
- [31] J. Racko, M. Mikolášek, A. Grmanová, L. Harmatha, J. Breza, D. Donoval, F. Schwierz, R. Granzner, and K. Fröhlich, "Trap assisted tunnelling current in MIM Structures", Presented at Progress in Applied Surface, Interface and Thin Film Science 2009.
- [32] R. Granzner, M. Zentgraf, and F. Schwierz, "On the vertical design of cubic AlGa_N/Ga_N HEMTs", Proc. 54th IWK, pp. 205-206, TU Ilmenau, October 2009.
- [33] J. Racko, A. Grmanova, L. Harmatha, J. Breza, D. Donoval, F. Schwierz, R. Granzner, and K. Fröhlich, "Trap assisted tunneling current in MIM structures", Proc. Electronic Devices and Systems IMAPS CS, International Conference, pp. 111-116 (2009).
- [34] W. Jatal, K. Tonisch, R. Granzner, M. Kittler, U. Baumann, F. Schwierz, and J. Pezoldt, "2H-AlGa_N/Ga_N HEMTs on 3C-SiC(111)/Si(111) substrates", Presented at ICSCRM 2009.
- [35] J. Racko, L. Harmatha, F. Schwierz, R. Granzner, J. Breza, and K. Fröhlich, "Current transport in MIM structures", Abstract Proc. 32nd Int. Spring Seminar on Electronic Technology "Hetero System Integration – The Path to New Solutions in Modern Electronics", pp. 256-257 (2009).
- [36] R. Granzner, F. Schwierz, Chr. Schippel, H. Höhnemann, and V. Dudek, "Performance Improvements of SiGe HBTs by Optimizing the Collector and Emitter Design", Abstract Book ISTDM, pp. 189-190, 2008.
- [37] Chr. Schippel, F. Schwierz, and R. Granzner, "Comparison of the Effects of SiC Design and Ge Distribution on the f_T vs. BV_{CEO} Characteristics of SiGe HBTs", Abstract Book ISTDM, pp. 217-218, 2008.

Danksagung

Eine Arbeit wie diese ist natürlich nicht ohne die fachliche und moralische Unterstützung durch Kollegen, Freunde und die Familie möglich.

Einen herausragenden Anteil am Zustandekommen dieser Arbeit hat mein Betreuer und Arbeitsgruppenleiter PD Dr. Frank Schwierz. Seine freundschaftliche und kollegiale Art, die anregenden Diskussionen und besonders seine unendliche Geduld haben die Arbeit in der Arbeitsgruppe „RF and Nano Devices“ zu einem Vergnügen werden lassen. Entscheidend für das Gelingen der Arbeit waren seine inhaltlichen Anregungen und kritischen Anmerkungen.

Mein besonderer Dank gilt Dr. Vladimir Polyakov für die Monte-Carlo Simulationen zum Vergleich der Transportmodelle, die kritische Überprüfung der analytischen Modelle zum Quantum Confinement, für die Programmierung des 1D Monte-Carlo Transportsimulators, für die erfolgreiche Zusammenarbeit an verschiedenen gemeinsamen Veröffentlichungen und, sehr wichtig für mich, für seine besondere Art auch die kompliziertesten physikalischen Zusammenhänge verständlich zu machen.

Bedanken möchte ich mich bei den aktuellen und ehemaligen Mitarbeitern des Instituts für Halbleitertechnik der RWTH Aachen sowie der AMO GmbH Aachen, insbesondere bei Prof. H. Kurz und Dr. Max Lemme für die erfolgreiche Zusammenarbeit in gemeinsamen Projekten zu Nanowire MOSFETs, die Prozessierung der Bauelemente und die gemeinsamen Publikationen.

Bei meinen Kollegen Christian Schippel, Dr. Mario Kittler und Jörg Gessner bedanke ich mich für die angenehme und fruchtbare Zusammenarbeit in verschiedenen Projekten und bei gemeinsamen Veröffentlichungen und, nicht zu vergessen, für die Hilfe bei Computerproblemen aller Art.

Stefan Thiele danke ich für die zahlreichen anregenden Diskussionen über Probleme der Festkörperphysik (Quantenkapazität, Phononen, Graphen) die letztlich zu einer gemeinsamen Veröffentlichung geführt haben.

Ein spezieller Dank geht an Dr. Toufik Sadi für die immer fruchtbaren und kurzweiligen (Fach-)Gespräche zu den verschiedensten Themen (numerische Simulation, Kompaktmodelle, Fußball, das Leben).

Den Mitarbeitern des Fachgebiets Festkörperelektronik danke ich für das angenehme Arbeitsklima.

Zum Schluss möchte ich mich bei meiner Familie, ganz besonders bei meinen Eltern für ihre uneingeschränkte Unterstützung und Verlässlichkeit in allen Lebenslagen bedanken.